ESSAYS ON CHINESE ENERGY POLICY AND MANUFACTURING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Thomas Eisenberg

August 2019

ESSAYS ON CHINESE ENERGY POLICY AND MANUFACTURING

Thomas Eisenberg, Ph.D.

Cornell University 2019

These essays explore sources of inefficiency and mismeasurement, as well as possible routes for improvement via policy, in China's coal power market and manufacturing sector. In the first chapter I develop a new model to quantify the behavior of market planners in China's coal power sector as well as how this behavior affects the investment decisions of power plants. Chapter 2 compares different approaches to measuring and diagnosing the possible sources of inefficiency in China's coal power market. Chapter 3 takes a broader look at the manufacturing sector in China to examine how differential treatment of state-owned enterprises should be accounted for in estimates of aggregate productivity for China.

BIOGRAPHICAL SKETCH

The author was born in Ithaca, NY and lived there until pursuing his undergraduate degree in Economics and Statistics at Swarthmore College in 2007. From there, he worked for 2 years in economic consulting at Charles River Associates in Washington, DC before coming back to Ithaca for a doctorate in Economics at Cornell. After graduation, he will start work as an Assistant Professor of Economics at the Lerner College of Business and Economics at the University of Delaware.

This document is dedicated to my late father, Theodore Eisenberg, without whom I never would have gone to graduate school. I am sure he would have had extensive and enthusiastic feedback for me on these papers.

ACKNOWLEDGEMENTS

First and foremost, I acknowledge the help of my committee: Panle Jia Barwick, Jean-Francois Houde, Shanjun Li, and Julieta Caunedo. These essays would be nowhere near the quality they are without all of their help, and every member helped to make my graduate school experience both rigorous and enjoyable.

Equally as important has been the help of my family, especially my mother, Lisa Eisenberg, who I got to see nearly every week in graduate school. My sisters, Kate and Annie Eisenberg, have also been important supporters of my research, as has my brother in law Aaron Blumkin. My family's two dogs, Barkley and Bertie, were an important source of moral and emotional support. I also extend a thank you to my niece and nephew Dylan and Oliver Blumkin, who are now both much taller than when I started graduate school.

I owe a special thanks to Ron Ehrenberg, who helped me get my start in economics research over 10 years ago, and who has remained a source of support and encouragement the entire time.

My work also benefited from many conversations with Francesca Molinari, who served on my committee earlier in graduate school. Todd Gerarden, Ivan Rudik, Giulia Brancaccio, Doug McKee, George Boyer, Marty Wells, and many others also helped, often as supportive and understanding bosses.

Many graduate school peers have helped advance my research, including Debi Prasad Mohapatra, Flavio Stanchi, Miriam Larson-Koester, Yang Zhang, Lin Xu, Caroline Walker, Andrew Fieldhouse, Barton Willage, and Maura Coughlin. I would also like to thank the many brilliant friends and organizers I met through Cornell Graduate Students United, and wish them luck on their future efforts. My cohortmates and friends on Uris 4th also helped me get through what was sometimes a long six years.

This dissertation would not have happened without the many mentors who encouraged me beforehand, including Adriana de la Huerta at Charles River Associates, and Amanda Bayer, David Huffman, and John Caskey at Swarthmore College. I would also like to thank the Center on Budget and Policy Priorities, where a brief internship helped hone my interest in economic policy.

Finally, I would like to acknowledge the generous financial support I received from the Hu Shih Fellowship in Chinese Studies through Cornell, whose funding helped bring this project to its full potential.

TABLE OF CONTENTS

CHAPTER 1

REGULATORY DISTORTIONS AND CAPACITY INVESTMENT: THE CASE OF CHINA'S COAL POWER INDUSTRY

1.1 Introduction

There is a large and growing literature on factor misallocation in developing countries, with a particular focus on China due to papers like Hsieh and Klenow (2009). China is a rapidly growing economy with varied and complex industrial policies across different market, which have recently received empirical attention (for example, Kalouptsidi (2018) and Jia Barwick et al. (2019) in shipbuilding). China's coal power market is the largest electricity market in the world, and China is especially dependent on coal power compared to the US and Western Europe, with over 70% of its electricity production coming from coal (Aden et al., 2009). Coal use also carries with it substantial environmental externalities from carbon emissions. This market is thus the most important market globally from the perspective of climate change policy. Any policy changes to this market carry large direct and indirect effects.

A common policy proposal is to "restructure" an electricity generation market. This would move production from planned, government-determined schedules, like China's current setup, to a market-based allocation scheme meant to promote efficiency and competition. For example, US wholesale electricity markets use multi-unit auctions that are on average extremely competitive. Such conversions have been studied extensively in the US and Europe by Fabrizio et al. (2007) and Newbery and Pollitt (1997). Gao and Van Biesebroeck (2014a) have analyzed initial reforms in the Chinese market in particular. These studies tend to focus on whether plants became more efficient on the intensive margin–that is, whether new incentives induced by market-based allocation encouraged plants to run more efficiently from a technical perspective.

Studies of electricity restructuring rarely answer a related question: given the existing physical infrastructure of a country's electrical grid, are planners assigning production in an efficient manner? This creates a form of output misallocation, where by having plants produce at differing levels, aggregate targets could be met at lower costs. Output misallocation in China's case can be thought of as resulting largely from favoritism shown to (or not shown to) different firms.

This favoritism comes with strong dynamic implications: if plants anticipate that they will not be awarded production in line with becoming more efficient and/or larger, they will change their investments capacity in response. This can lead to long-term welfare losses, as investments in power generation capacity are usually irreversible, and even in the aftermath of policy intervention welfare losses may persist. Studies seldom address these dynamic implications, which are especially important in a rapidly growing economy like China's.

China presents a uniquely challenging regulatory environment for analysis: decisions are made by differing authorities across markets, policies may change exogenously and without warning from central planners, and growth has been so rapid in recent years that the market is likely not governed by a stable, stationary dynamic process. As such, a uniquely rich dynamic model that accounts for the complexity of this market is necessary. Quantifying plant-level investment policy responses is especially important in China, since during my period of analysis demand was growing especially fast and both new and old plants sought to expand rapidly to meet newly higher production targets.

This paper has two goals: to estimate the extent of output misallocation generated by planning policies in China's coal power industry, and to quantify the effect of these policies on plant-level capacity investment. The answers to these questions help to measure the potential short- and long-term effects on plants from policies like market restructuring, and they also shed light on the current objectives of planners in China. Planners, who are solely responsible for allocating production across power plants, face many tradeoffs in making their decisions, including political considerations, notions of fairness and market concentration, and labor concerns. The major contribution of this paper is to develop a framework that measures the net effect of all of these considerations, and model the dynamic investment decisions of plants who take them as a primary factor in determining their production. This paper focuses solely on the generation side of China's coal power sector, as opposed to the transmission or distribution side.

The main data set for this analysis is a combination of the commonly used Annual Survey of Manufacturers from China's National Bureau of Statistics, and a confidential census of coal power plants conducted by the Chinese government called China Power Plant Annual Statistics. This allows for the construction of a novel dataset which includes output price, input price, electricity production, coal use, nameplate capacity, and heat rate information for over 1500 coal power plants. I supplement this with province-level electricity import and export data from the proprietary CEIC database. For high-level comparisons to the United States, I use plant-level data from the US Energy Information Administration.

The empirical framework of this paper consists of two stages: a static model and a dynamic model. I first develop and implement a novel method to identify costly losses from misallocated production in China's coal power market. In turn, I develop a tractable dynamic model of this industry to gauge the effects these distortions have on a plant's capacity investment decision, and the implications this may have for costs and market shares. Electricity production often involves a dynamic investment decision, but plants under a policy regime like China's face an additional consideration: shocks to the amount of electricity they are allowed to produce may be persistent or even time invariant, which may drastically distort their lifetime expected stream of profits.

The key objects of interest from the static model can be thought of as output "wedges" from the misallocation literature (see, ie Restuccia and Rogerson (2008) or Hsieh and Klenow (2009)). These wedges rationalize a plant's observed production versus what it would be given under "efficient dispatch", where plants are allocated production based purely on their cost ranking (and random shocks). My approach is exploits the planned nature of this market, as I can explicitly model the allocation of output by regulators. Planners make discrete choices that depend on a plant's capacity, marginal cost, and wedge (or policy distortion). Each wedge can be thought of representing the effect of unobserved constraints or inputs into each planner's objective function.

Despite the findings in Gao and Van Biesebroeck (2014a) that initial reforms were successful, it is widely understood that China has yet to truly move away from planned production in this market, and according to Resources for the Future, there still is "no spot market" for coal power generation in China (Ho et al., 2017). The wedges can also be thought of as capturing the net effect of multiple levels of industrial policy in this market, similar in spirit to structural papers like Kalouptsidi (2018) or reduced form papers like Lane (2017).

Given an estimated wedge for each plant-year, I then specify and estimate a dynamic discrete choice model of nameplate capacity investment where plants track their costs, capacity, and wedges. Taking a plant's decision to enter as exogenous, I also model a plant's initial period investment decision. Investment is particularly important in the power generation context, as a plant's ability to generate electricity is strictly limited by its nameplate capacity. Additionally, private investment is both allowed and encouraged, so the decision to invest is essentially a profit-maximizing one, subject to the onerous regulatory framework of the static market.

While there have been many recent empirical studies that look at electricity in developing countries, and possible efficiency losses from government policy (such as Ryan (2014) or Gao and Van Biesebroeck (2014a)), as well many empirical dynamic papers on power plants (like Abito (2017)), very few studies combine the two. This paper contributes to that literature while examining a policy-relevant and complex setting like China.

China's coal power industry presents a uniquely difficult dynamic context for several reasons. First, market demand (and thus aggregate capacity and supply) are growing rapidly through my entire sample period of 1998-2007. This suggests that the dynamic environment plants face is extremely nonstationary. Second, there are several large one-time quasi-exogenous policy changes that may drastically affect payoffs in this market. Chief among these are the 2002 restructuring reforms which plausibly changed the structure of almost every province, and which elicited plant-level responses in efficiency according to Gao and Van Biesebroeck (2014a). Third, there are a second set of policy changes that may differ across markets, such as a brief experimentation with market-based mechanisms in Guangdong, that mean returns on investment may differ significantly across provinces.

These three facts form the basis of my dynamic model. Starting with a backbone of a three-dimensional continuous-state Rust (1987) style model, I incorporate a combination of perfect foresight assumptions from the durable goods literature like Conlon (2012), nonstationary aggregate state approximation models from Weintraub et al. (2008) and Weintraub et al. (2017), nonstationary nested fixed point methods from Rust and Phelan (1997), and parallel computing methods to quickly calculate value functions that vary across markets. Overall, the dynamic analysis of this paper borrows methodologically from a wide range of sources. I also incorporate approximation methods for state space reduction similar to those in Leslie and Sorensen (2014), and solve everything using collocation methods as describe in Fackler and Miranda (2002). To my knowledge, this is the first analysis that estimates a capacity investment model with all of these features.

This combination of methods allows me to make key simplifications: like in the agricultural context of Scott (2013), there is a large universe of possible aggregate states (average prices, total demand, total market capacity) that could influence the government's allocation policy, and thus each plant's expected returns. Additionally, there is a large amount of missing data for several candidate aggregate states, like average output price. Assuming plants take the year to be a state variable and have perfect foresight over its sequence, similar to Conlon (2012), provides a computationally tractable way to circumvent these issues. Plants are aware of the states that influence policy, and they only care about their evolution to the extent that it affects their payoffs. Thus, year and market specific value functions should capture all possible payoff-relevant information that plants are exploiting. These methods come with a cost: I am unable to reconstruct full counterfactual equilibria that incorporate plants responding to each other, and have to focus any counterfactual analysis on representative firms (done in many similar papers, including Timmins (2002)).

I find reduced form evidence that suggests unobserved heterogeneity in investment costs is necessary to include in any structural dynamic model. Power plant data provides relatively few observable variables that are informative in this dimension beyond costs and capacity, so I use the EM algorithm to estimate a mixture model of discrete investment types, similar to Keane and Wolpin (1994), Scott (2013). This comes with two problems: First, mixture models are generally improved if it is possible to specify a relationship between an observation's initial conditions and its type. Second, absent this, my estimates tend toward a corner solution where all plants that never invest are grouped into one "type" where investment costs are unrealistically high. Modeling the initial investment decision of entering plants largely resolves both issues. Importantly, this also allows a plant's type to be correlated with its initial capacity level.

The first-stage static exercise measures an upper bound on the gains from market restructuring. I find that, on average, provinces would be able to meet aggregate electricity demand at a 2.8% lower per-unit cost if planners fully prioritized marginal costs in assigning production. Within my sample, this corresponds to roughly $3.8 billion in savings. I find there would be an additional $93 - $900 million in savings from carbon emissions. These savings stem largely from smaller, higher-cost plants being favored by current allocation policies. Kahrl et al. (2013) investigate detailed engineering data for one province and find optimal cost-based dispatch would generate similar savings of 4-5%. I find removing the policy distortions brings correlations between utilization and marginal cost measures in China much closer to those in the US. This evidence suggests removing the wedges is roughly equivalent to a practical upper bound for the gains from moving to a market-based system.

In the dynamic setting, I find several key interactions between output misallocation and investment: First, plants that do not forecast their expected distortions change their investment behavior by less than one third the amount that a forward-looking plant does. Second, investment behavior is extremely sensitive to a plant's current and future wedges, with a persistent 1 standard deviation reduction resulting in a 25% decrease in a plant's probability of making a large investment. Higher wedges increase a plant's investment frequency more than an equivalent decrease reduces it. Third, entrant investment decisions are even more sensitive to a plant's anticipated production stream.

Counterfactual simulated investment paths for plants at different levels of wedges show that, because of the cost savings associated with scaling up in power plants, costs may differ substantially after a short period of time. On average, after only 7 simulated years, per-unit costs for a plant with identical initial conditions differ by over 1.5%.

Additional simulations help to uncover some of the potential objectives that regulators are pursuing in this market. The evidence is clear that there are efficiency losses from output misallocation in this market, and that the distorted investment incentives likely amplify them. This leads to the question: what could China be gaining from the current allocation regime? One potential target is market concentration. My findings suggest that current allocation policies have a significant influence in this regard. I estimate that, starting in 2000, by 2007 a 75th percentile plant is on average 1.67 times larger than a 25th percentile plant after investing under observed production allocation streams. If these allocations are "swapped", that is, the 75th percentile plant receives the expected stream of the 25th percentile plant and vice versa, this number jumps to 2.45.

Observed policies are thus consistent with an effort to keep concentration low. This finding mirrors many of those found in papers on preferential treatment in procurement auctions, such as Saini (2011), Krasnokutskata and Seim (2011), and Marion (2011). This is by no means the only additional potential concern in planners' policy functions, with other possible inputs including labor concerns, promoting aggregate capacity, and avoiding infrastructure constraints.

The history and explicit policies of this market support the limiting of concentration as a goal: in 2002, the Chinese government broke up a large state-owned generation company with 50% market share into 5 smaller companies. But, they did not remove the strict price and quantity controls that power plants face in this market. Similarly, while provincial authorities have a fair amount of freedom in conducting their electricity dispatch, there is a general "guiding principle" of "fair dispatch" according to Ho et al. (2017) and others, where many planners are explicitly trying to equalize utilizations across plants. The exercise of market power is a common concern in restructured electricity markets, discussed extensively in articles like Borenstein et al. (1999) and Borenstein et al. (2002).

This paper draws from recent methodological advances in dynamic estimation, and while I do not explicitly model strategic interactions between firms, the recent literature on this subject such as Bajari et al. (2007), Aguirregabiria and Mira (2007), and Sweeting (2013) relates closely to this work.

In summary, my empirical framework allows me to recover both policy distortions and a plant's investment policy function in an extremely complex environment, while remaining faithful to the regulatory environment that plants face in China.

In section 2, I further discuss the industry background, context, and data sources. In section 3, I present motivating reduced form evidence. In section 4, I present both a static model of planners' decisions to allocate production and a dynamic discrete choice model of capacity investment. Section 5 discusses estimation and parameter estimates, while Section 6 discusses counterfactual results. Section 7 concludes.

1.2 Industry Background and Data Description

1.2.1 Industry Background

China's current electricity production model began around 1998. according to Xu and Chen (2006), 1998 marked the start of a shift toward efficiency-focused reforms after decades of growth-focused policy. The state no longer held a total monopoly over the power generation industry like it traditionally had, and the market now had "a...structure composed of diversified investors" (Xu and Chen, 2006). However, this by no means resulted in a smoothly-functioning market system: "The reform in the electricity industry was mainly on the governmental level, the old regulatory system did not change at all in the lower levels, which remained incompatible with both the power industry's market-oriented reform and diversified operating entities...influence from the central government was still very large and the governments, both central and regional, played an important role in the industry. A modern regulatory system was far from coming into being" (Xu and Chen, 2006).

Put differently, the central, regional, and provincial governments all still played (sometimes conflicting) roles in a plant's operation. These actors often had differing political and economic objectives: a provincial head would likely care about maximizing province-level output or profits rather than ensuring a more efficient allocation of resources across a wider geographical areas. This is especially important in China, where coal resources are not evenly distributed across the country. Xu and Chen (2006) state: "Areas rich in primary energy deposits were far from power-load centers. However, market segmentation by administrative divisions exerted a tremendous impact upon resource allocation; power from cheap, clean energy sources were rarely distributed across provincial divides due to inter-political barriers." Only adding to these frictions is China's underdeveloped transmission apparatus, which adds a physical barrier to the existing political ones.

Xu and Chen (2006) also indict the pricing system, claiming there was both a lack of uniformity and enforcement across plants. As a result, "prices could not reflect the true relationship between supply and demand." The authors also claim that pricing and investment conditions placed independent power producers "at a disadvantage compared to state power plants" at times when there was enough capacity to meet demand.

In 2002, there was a major attempt to reform some of these processes, with regulatory authorities breaking up the largest state owned plant into five smaller ones. Gao and Van Biesebroeck (2014a) estimate the effects of these reforms and find that there were, in fact, modest efficiency gains, in that some plants became more technically efficient in response. Despite these gains, it is widely acknowledged that the "second stage" of the reforms never took place: power plants were never allowed to independently set prices or quantities, and it was widely agreed that small-scale experiments in creating true markets did not produce the returns that the Chinese government wanted.

Limited competition was first introduced experimentally in 2000 in Shanghai, Zhejiang and Shandong, and later that year in Liaoning, Jilin and Heliongjiang (Ma, 2011). Within these provinces, only a subset of plants were chosen, and only small portions of their production were eligible for the programs. These experimental reforms contributed to large differences across markets in addition to their already substantially different underlying qualities like density or natural resource endowments. While efforts have been made to separate China into 6 regional grids that can freely transmit electricity between provinces have been made, production and pricing decisions still mostly take place at the provincial level (see, ie, Ho et al. (2017) and Liu et al. (2013)).

2002 and 2003 also saw drastic changes in coal input markets. Coal had generally been sold on a "two-track" system prior to these years, where quotas of coal were sold at fully planned prices so as to ensure electricity production was met at certain cost targets (see Ma (2011)). After these quotas were met, all remaining coal was sold via market mechanisms. Coal input markets began to transition to fully market-based allocation schemes around this time. As documented in Liu et al. (2013), this lead to an increase in coal prices for most power plants, and power plants were unable to similarly raise their output prices due to policy constraints. A limited policy response later followed: in both 2004 and 2006, the government instituted "coal and electricity price co-move" policies, that allowed for limited windows of price increases.

Taken in total, the various market-specific and national policy changes lead to an extremely complex and volatile dynamic environment for plants. Any plant that is forecasting its lifetime profits in order to make an optimal investment decision would be accounting for the extreme non-stationarity of this market as a first-order concern. Any dynamic model of this market to be able to capture the extreme heterogeneity in payoffs that plants may face (and thus forecast) across each year and market.

My sample extends through 2007, and all of these policy examples occur during my sample period. However, the literature suggests that many of these cost and efficiency problems persist well after 2002: Liu et al. (2013), writing in 2013, say that "power-generating companies...must sell their output at regulated prices that often do not cover costs." In the present day this results in some power plants cutting their supply in protest, resulting in painful blackouts for many people. Thus, while my sample period is largely historical, the question of how to reform allocation mechanisms in this market remains a pressing policy question.

Smaller scale papers on this market have provided some initial answers: Zhao and

Ma (2013) study a panel of 34 large power plants and find that there have been efficiency gains in response to the 2002 restructuring. Ma (2011) provides an overview and empirical summary of the "on-grid tariffs", or prices that plants have been allowed to charge for their output, over the past couple of decades. Liu et al. (2013) establish an empirical link between coal prices and electricity output prices, and find that electricity pricing authorities are generally slow to respond to market-driven changes in the coal price.

Investment

Ren et al. (2019) detail how China has ambitiously invested in coal power despite attempts to move to more environmentally-friendly field sources. While their analysis focuses mainly on later years than this paper, 2013-2016, it provides some key context for earlier periods. Namely, explain that investment approval processes operated under relatively centralized mechanisms until 2014, at which point control shifted almost entirely to province-level regulators.

Under this centralized regime, the government explicitly directed that coal plants be given close to equal utilizations determined by annual contracts so as to encourage investment. Prior to 2003, prices still did not operate in such a way that investment was especially profitable, while market-oriented reforms around 2002 and 2003 began to relax the strict pricing standards that had previously been in place.

The literal investment process has historically been very opaque. According to Ren et al. (2019): "Prior to 2014, the central government retained sole authority to approve coal-fired power projects. The approval process was often lengthy and costly..." The authors also mention that there is a lack of transparent criteria for investment approval, and that while central authorities approve investment, provincial governments are often involved in the process via their role as investors in the power plants.

Taken together, the relationship between investment and production in Chinese coal power can be summarized in the following way: investment is often incentivized through granting plants favorable production or prices, and due to opaque approval processes this is a necessary "carrot" to provide to some plants. Power plants generally are subject to similar approval rules across provinces, and applications for investment are often made by private (or pseudo-private) parties.

1.2.2 Data Sources

The key dataset to this paper is a confidential survey of coal power plants conducted by the Chinese government. It covers, roughly, the universe of power plants from 1995, 1997-1998, 2000, and 2002-2011 [1]. Major variables include a plant's name, power generated, coal used, and nameplate capacity. The plant's name allows us to find locations and ownership status–the latter is extremely important for determining which plants were and are owned by the "big 5" state-run corporations, as well as plants that are owned partially by the state. The fullest version of the dataset contains 21,121 plant-year observations. From this data I can also derive a plant's "heat rate", a standard measure of efficiency calculated by dividing coal input by power output. This will be the main index I use to assess cross-plant physical efficiency levels, and the associated emissions from each plant's output in counterfactual scenarios.

My analysis primarily depends on a subsample of plants that can be merged with

[1]Thank you to Shanjun Li, Deyu Rao, and many others for preparing this data and allowing me to access it.

the annual manufacturing survey from the Chinese National Bureau of Statistics to get revenue and expenditure information. I refer to this as the "revenue" sample. I use the larger sample with more limited variables to aid in the construction of competitive fringes and market-level aggregates in my analysis. A smaller subset of observations (mostly determined by markets that are populated enough) is then fed into the dynamic analysis (the "dynamic" sample). See Appendix for the full details and comparisons of each sample. Compared to the physical sample, roughly 60% of aggregate capacity survives when financial variables are included. This final sample includes 504 unique plants.

Table 1.1: Means of Major Variables, 1998-2007

Year	Cap (MW)	Prod (MW)	Price (000 RMB/MWh)	Phys. Cost (000 RMB/MWh)	Heat Rate (tons/MWh)	N
1998	384.58	216.73	0.26	0.17	0.55	195
2000	422.75	239.15	0.28	0.17	0.60	193
2002	484.50	297.51	0.27	0.17	0.60	221
2003	491.99	328.71	0.29	0.19	0.55	232
2004	509.97	349.23	0.26	0.19	0.56	283
2005	549.39	361.71	0.29	0.23	0.56	292
2006	636.19	391.38	0.28	0.22	0.54	326
2007	693.14	416.70	0.31	0.25	0.53	351

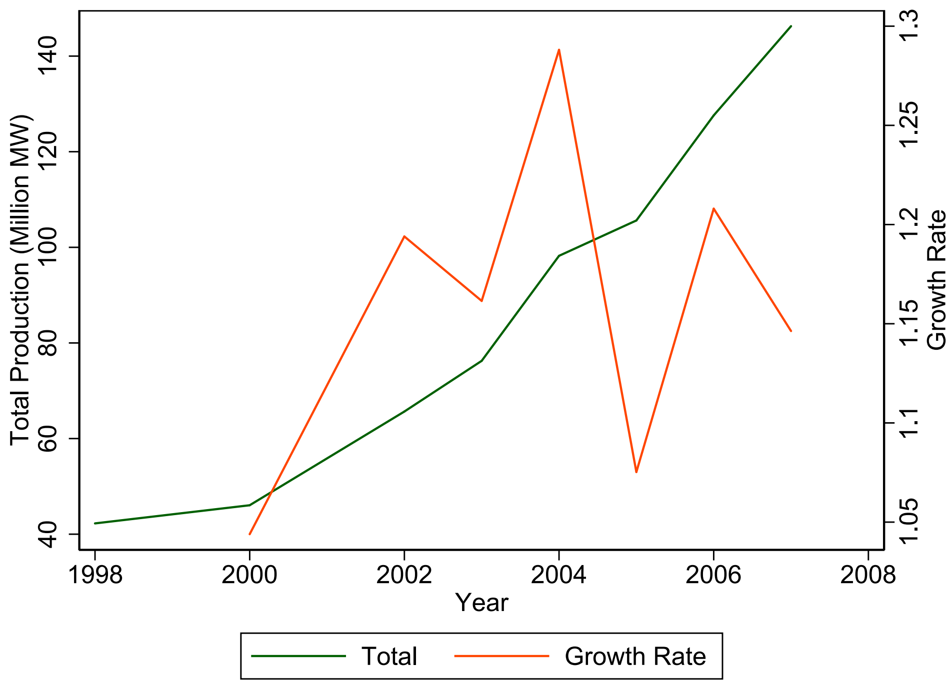
Notes: Table depicts summary statistics for years 1998-2007. Physical variables are from confidential power plant survey, financial variables are from a combination of physical dataset and financial variables from annual NBS manufacturing census. One RMB is roughly .15 dollars, so the output price in 1998 of .26 000 RMB/MWh would equal about 40 dollars per MWh, while the 2007 output price would be more like 47 dollars. Figures are for "revenue" sample.

This analysis also makes use of electricity trade balance data across provinces in China. The CEIC database for China lists electricity imports and exports at the provincial level for 1995-2014. This allows me to construct a trade balance for each province-year combination, so the structural model can at least control for situations like Beijing or Tianjin, which are essentially city states that must import most of their electricity. Conversely, there are western provinces like Gansu that export a high percentage of their production since they are sparsely populated but rich with coal resources.

1.2.3 Trends

Even from this fairly limited sample, it is plain to see the nonstationary environment plants are operating in:

Table 1.2: Total Electricity Production and Growth Rate By Year, 1998-2007

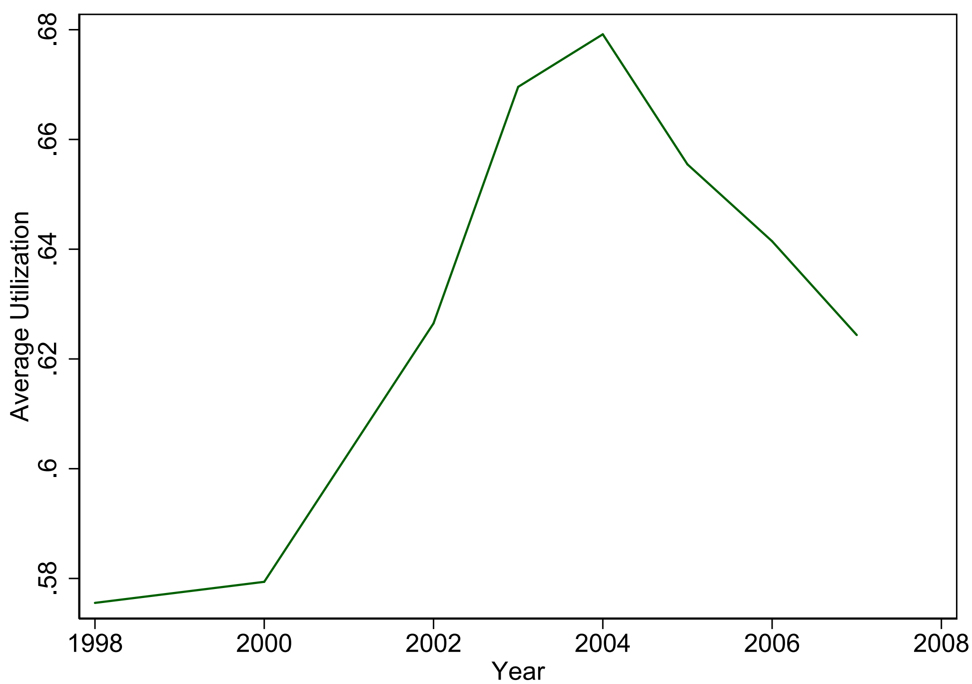


Notes: Restricted only to sample that has full set of physical and financial variables. For missing years, growth is assumed to be constant across the 2 year gap. Production data comes from confidential power plant survey.

Production grows rapidly each year, and is extremely volatile. From 2003 to 2004, production multiplies by almost 1.3. This is consistent both with the large amounts of net entry seen into the market, as well as demand and production potentially growing for incumbent plants. A forward-looking plant could forecast aggregate production or demand as a state variable, but it is difficult to say directly how it should be affecting their payoffs. One way to more directly examine how plant-level production changes is to look at aggregate utilization. If this measure shifts, this suggests that the changing market-level conditions are translating into individual payoff changes, rather than a situation where investment and entry just keep plants at the same level relative to aggregate demand.

Utilizations exhibit a somewhat different pattern from production–capacity starts to get strained by about 2004, but then utilization decreases. Demand is growing during this period, so this loosening of capacity constraints must be coming from investment and/or entry:

Figure 1.1: Average Utilization By Year, 1998-2007

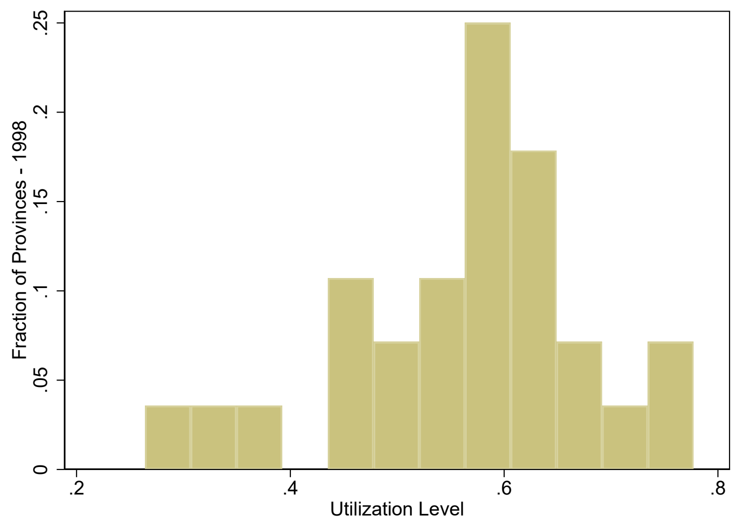


Notes: Restricted only to sample that has full set of physical and financial variables. For missing years, growth is assumed to be constant across the 2 year gap. Production and capacity data comes from confidential power plant survey.

Average utilization does not follow a clear linear trend through time. Not only does this environment appear to be non-stationary, but key candidate aggregate states (demand, aggregate capacity, average utilization) do not abide by an obvious common pattern either. Thus, dealing with non-stationarity by choosing a specific aggregate process to model that plants can forecast is not straightforward.

Finally, it is clear that policy and state-variable differences across provinces also generate very different aggregate outcomes:

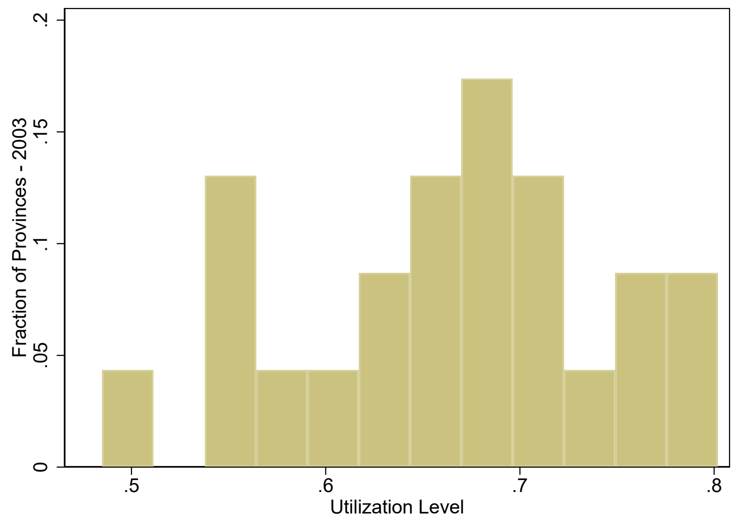
Figure 1.2: Histogram of Province-Level Utilizations, 1998



Notes: Restricted only to sample that has full set of physical and financial variables. Production and capacity data come from confidential power plant survey. Includes 30 provinces.

With some provinces averaging as low as .2 utilization overall in 1998, and others as high as above .7, heterogeneity across provinces is consequential to plants. Given these differing levels of production, the return on investment already differs across markets at the earliest point in the sample.

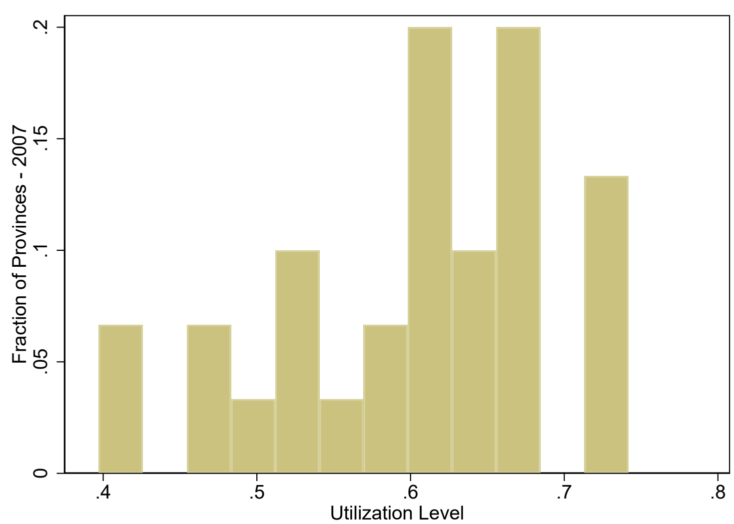
Figure 1.3: Histogram of Province-Level Utilizations, 2003



Notes: Restricted only to sample that has full set of physical and financial variables. Production and capacity data come from confidential power plant survey. Includes 30 provinces.

As of 2003, average province-level utilizations have clearly shifted higher, but the distribution is also less peaked now. Thus, the differing initial utilizations seem to be on different trajectories as well.

Figure 1.4: Histogram of Province-Level Utilizations, 2007

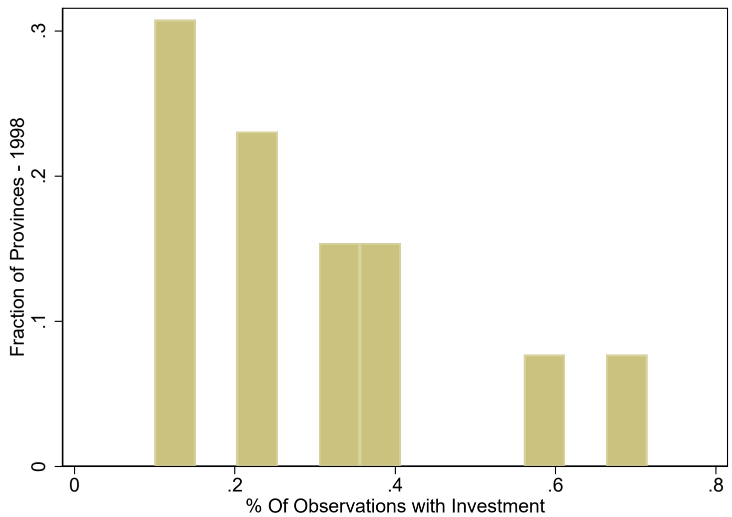


Notes: Restricted only to sample that has full set of physical and financial variables. Production and capacity data come from confidential power plant survey. Includes 30 provinces.

2007 provides yet another permutation: a bimodal distribution between .6 and .7, no province above .75, and no province below .4 like in 1998. Provinces are neither staying at the same level over time, nor are they converging to a common level.

This provincial heterogeneity translates into differing investment rates:

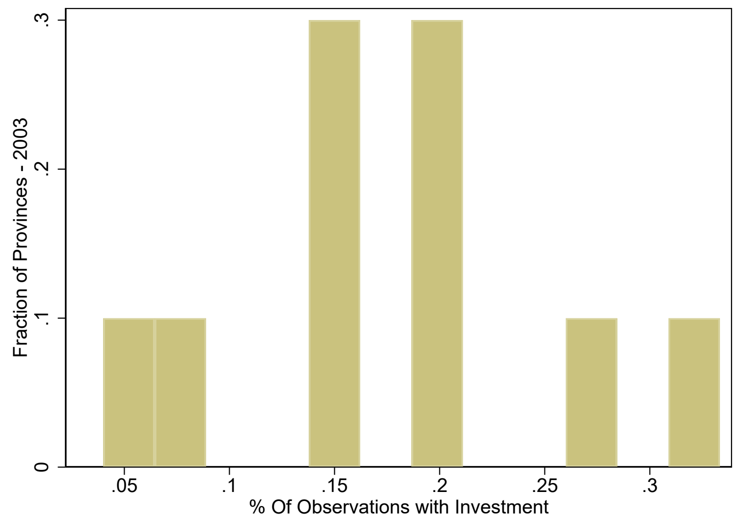
Figure 1.5: Investment Frequencies by Province, 1998



Notes: Requires investment to be at least 20% of existing capacity. Firms are assumed to make only one investment over two year periods generated by missing data. Graph covers all (30) provinces. Data comes from capacity panel in confidential physical power plant survey. Includes 30 provinces.

The modal investment frequencies in 1998 appear to be around 10 and 20%. There is enough heterogeneity to suggest different provinces may have systematically different investment policies, and that any analysis that combines their data will not be identified solely off of the investments in one or two provinces.

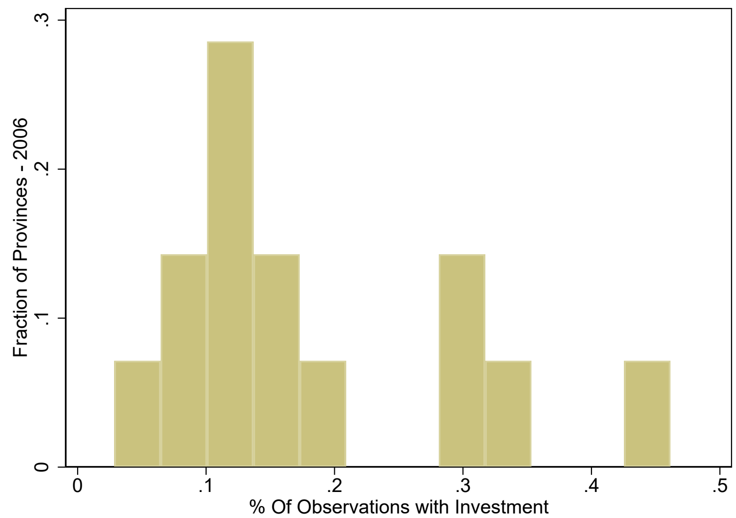
Figure 1.6: Investment Frequencies by Province, 2003



Notes: Requires investment to be at least 20% of existing capacity. Firms are assumed to make only one investment over two year periods generated by missing data. Graph covers all (30) provinces. Data comes from capacity panel in confidential physical power plant survey. Includes 30 provinces.

Investment rates are overall much lower in 2003 than in 1998. The number of plants is much smaller in 1998, and plants that invested back then are on average extremely unlikely to do so again. Overall, this is an extremely different distribution only 5 years later.

Figure 1.7: Investment Frequencies by Province, 2006

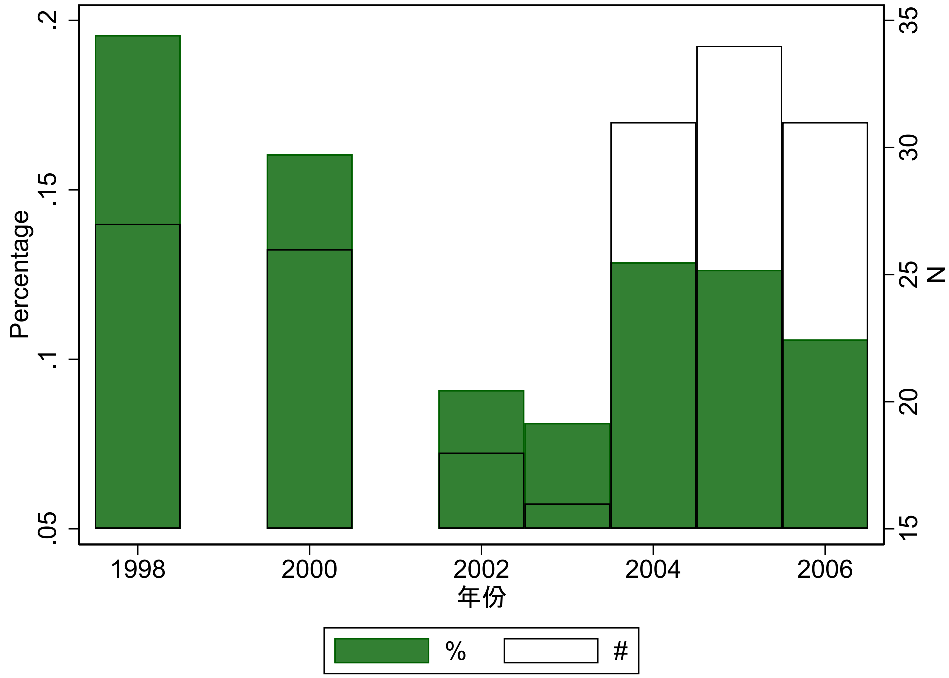


Notes: Requires investment to be at least 20% of existing capacity. Firms are assumed to make only one investment over two year periods generated by missing data. Graph covers all (30) provinces. Data comes from capacity panel in confidential physical power plant survey. Includes 30 provinces.

2007 is somewhere between 1998 and 2003 in terms of how many plants are investing. Rates are lower than at the start of the sample, and entry rates have likely slowed down by now. But, there are now a fair amount of relatively recent entrants who are still trying to reach efficient scale.

With the heterogeneous conditions established across time and markets, we can investigate investment rates more directly. It is important to look at who is investing, how much investment is taking place, how often plants are investing, and how investment and cost measures relate to get a sense of how plants are responding to their conditions. The following graph provides the percentage (within a given year) and number of plants that invest for each year investment data is available.

Figure 1.8: Percentage and Number of Firms Investing By Year

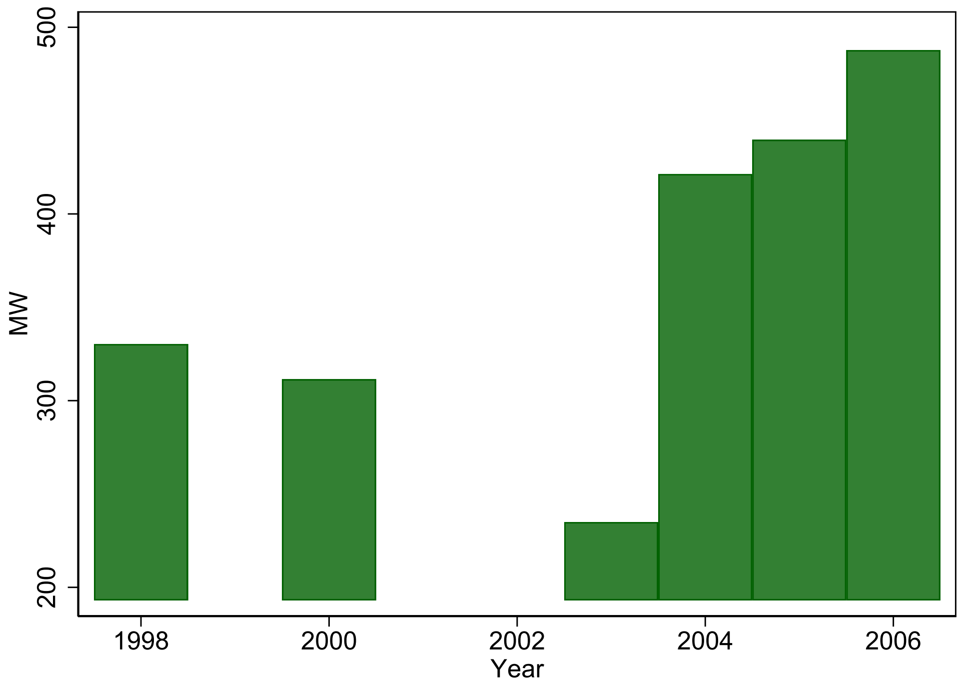


Notes: Requires investment to be at least 20% of existing capacity. 1999 and 2001 are missing. Graph covers all provinces. Data comes from capacity panel in confidential physical power plant survey.

For almost every year, at least 10% of plants invest. The percentage of investors falls sharply starting around 2006, but the raw number of investors is more stable. Part of this is mechanical: there is net entry into the sample over this span. It is also possible that the 2002 restructuring effort affected investment patterns, though this is difficult to entangle from this graph alone given the background entry and exit patterns happening. At the very least, investment was common, lumpy, and occurring across a range of plants. After the sample ends, the global financial crisis also likely slowed investments for a couple of years, while the overall Chinese "boom" was a bit slower as well.

Investment decisions have both an extensive margin (whether it is made or not) and an intensive margin. It may be that investment rates and investment sizes followed drastically different patterns throughout the sample:

Figure 1.9: Average Investment Size By Year (MW)

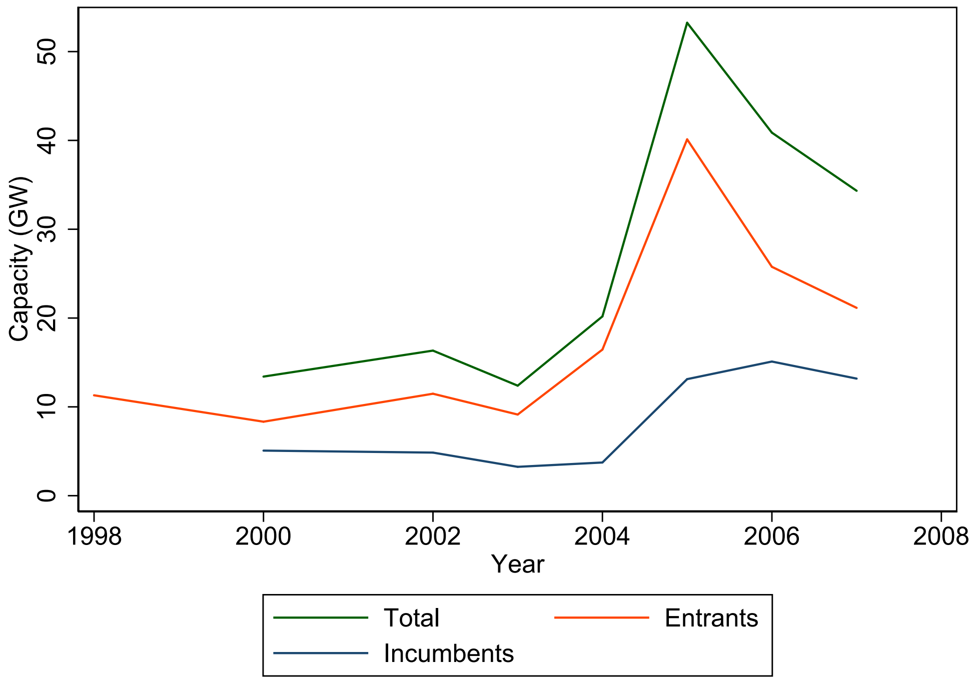


Notes: Requires investment to be at least 20% of existing capacity. 1999 and 2001 are missing. Graph covers all provinces. Data comes from capacity panel in confidential physical power plant survey.

This graph suggests that while investment sizes vary over the years, there is probably not any kind of dramatic technological or scale shift until possibly 2004. There is a brief decline in 2003, which may be explained by uncertainty in the wake of the 2002 restructuring. But, this is likely a change of degree rather than type, and that investments across years in this sample are probably directly comparable.

At this point we have established that investment and production patterns vary heavily by year and market, but it is also important to see how these translated into actual capacity growth. This is a period of substantial net entry in China, and both incumbent investment and capacity additions due to entrants are important in determining the environment plants face:

Figure 1.10: Incumbent vs. Entrant Capacity Added, 1998-2007



Notes: Added capacity is entrants in the current year plus investments made last year by incumbents. Graph covers all provinces. Data comes from capacity panel in confidential physical power plant survey. 1999 and 2001 are missing and linear trends are assumed for these years. Covers only firms for which financial and revenue information are both available.

After 2002 there is a spike in both entry sizes and investment. This reinforces the story that there are at least two "periods" firms are facing as they forecast.

Aggregate reduced form evidence has helped to establish several key facts: in the aggregate, both across years and markets, firms are facing a highly nonstationary environment with a large amount of policy variation. It is also clear that this is translating into different plant-level policy responses. To understand how the regulatory environment affects each plant individually, a more granular analysis of each investment decision is necessary.

1.2.4 Plant-Level Analysis

The basic story that this paper seeks to uncover is the following: there is a substantial amount of dispersion in marginal costs across power plants in this market, and given that production is set by central planners in China, dispatch could be done along efficiency-promoting lines. In reality, this does not seem to occur. Furthermore, plants may be forward looking and anticipate this decoupling of production and efficiency, and will change their investment behavior in response. We have seen in the prior section that this all takes place in a highly complex regulatory setting, and I now establish some key micro-level facts about costs and investments without imposing a structural model so as to ensure that more a complex analysis is not inventing results that do not match with simpler data patterns.

Costs

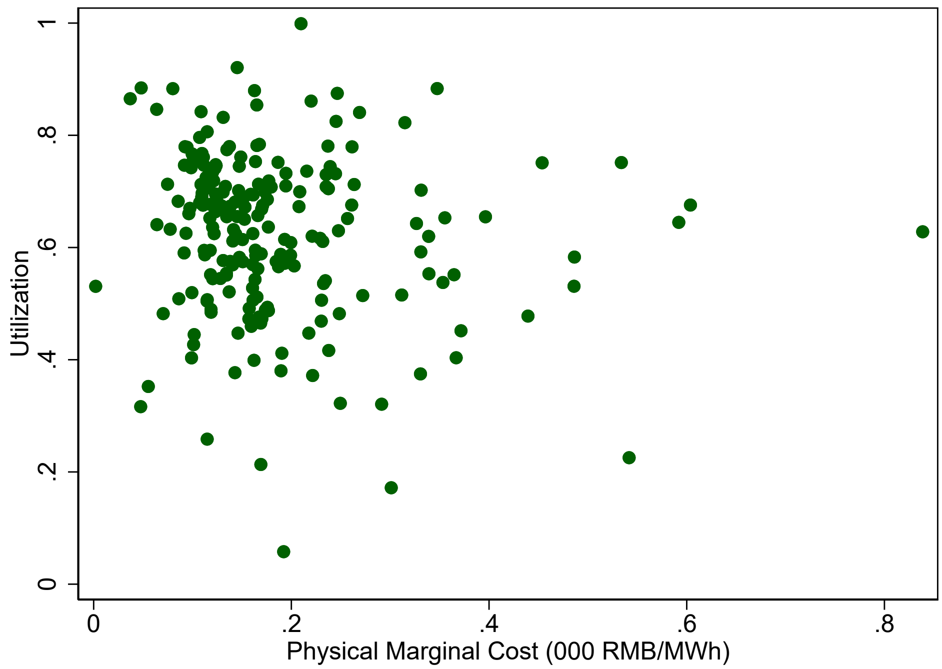
This analysis exploits the fact that once a plant is built, fuel costs account for a massive share of its operating costs. While investment costs tend to be between 30 and 50% of a plant's total costs, fuel costs represent 70 to 90% of the marginal costs that would be incurred for an already operating plant [2]. For a coal fired plant, coal is the near-exclusive source of fuel [3]. Misallocation in output thus maps directly to misallocation in coal. I identify misallocation via the insight that, accounting for nameplate capacity (a hard capacity constraint), a lower coal cost plant would be allocated more production (and thus coal) in an optimal planning regime. It is thus worth investigating patterns in

[2]For reference, in my dataset the median ratio of wage expenditures to input expenditures is between in 7 and 8%.

[3]In my dataset, oil is on average used .1% as much as coal, and the vast majority of plants record no gas use.

plant-level marginal coal costs, both to establish that dispersion in this measure exists, and to examine correlations between these costs and production:

Figure 1.11: Utilization vs. Cost: 2002



Notes: Includes data for all (30) provinces. Data are from confidential survey of powerplants and NBS annual manufacturing survey which provides input price information. One dot represents one powerplant. Note that costs and utilization are negatively correlated.

While 100% utilization in the lowest cost plant is likely unachievable for many reasons, we would expect to at the very least see a strong negative correlation between cost and utilization in a regime concerned with efficiency. The flat slope of the above graph is unsurprising, as there are many directives in place to equalize utilization across plants regardless of cost (see Liu et al. (2013)) in China. This is not an iron law of electricity production in China, but rather a "guiding principle."

This provides two key pieces of information: first, being able to plan anything like optimal cost dispatch in the face of hourly or daily cost shocks, infrastructure shortcomings, and other possible flaws is likely infeasible. There may also be room for improvement in reallocating production to Chinese coal power plants, but any model trying to assess this has to account for the complex and difficult nature of assigning dispatch under uncertainty. A simple comparison of current production modes to optimal dispatch based only on observed costs, while informative, will likely vastly overstate the possible gains.

The US provides a useful baseline for comparison:

Table 1.3: Correlation Between Utilization and Heat Rate: US vs China by Year

Year	China	US
2000	-.03	-.34
2002	.02	-.39
2003	-.13	-.15
2004	-.05	-.21
2005	-.02	-.20
2006	.09	-.46

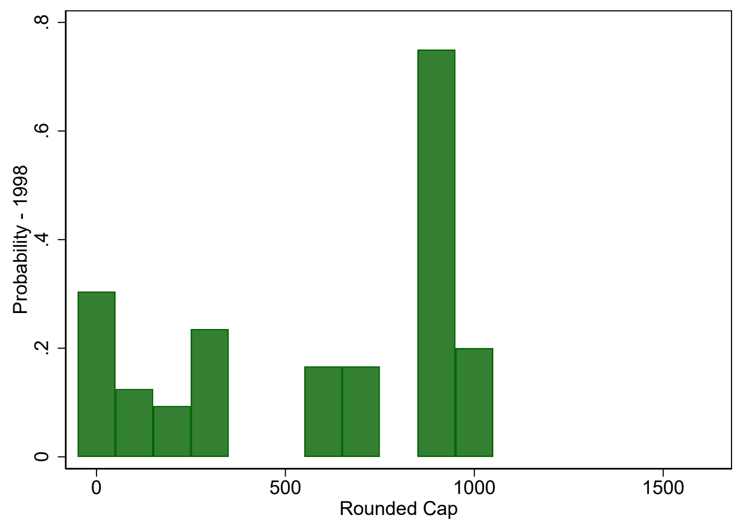
Notes: Table depicts correlations between heat rate and utilization in China and US from 2000 to 2006. Chinese data are from confidential power plant survey, and US data are from EIA. *N* for US is 3,694.

While these comparisons are in terms of heat rate rather than financial marginal costs, the two are obviously very highly correlated. Here we see that the Chinese correlations are much higher in every year, though this measure is somewhat noisy. If we take the US to be a measure of what is achievable in a restructured market, it is clear that the correlation between utilization and production generally does not approach -1. Logistic and physical constraints clearly play a large role in determining how dispatch can be allocated.

Investment

A key determinant of plant-level cost is its capacity. Ex ante, we would expect a higher investment probability for lower cost plants of the same size, as they expect to gain more profits over time

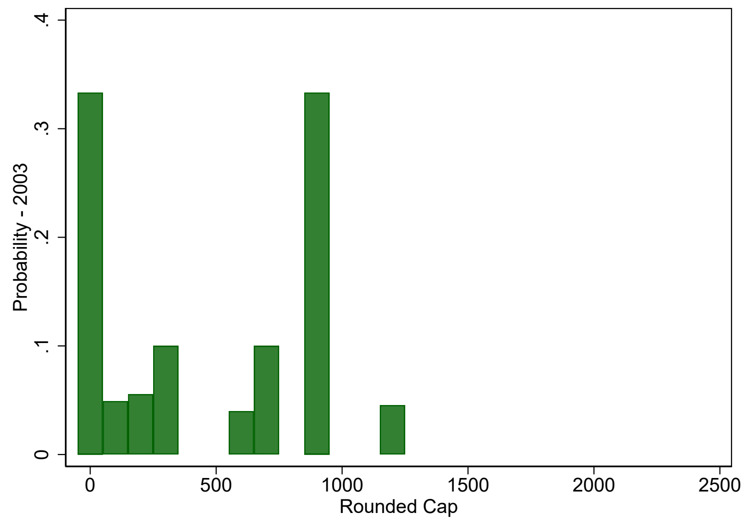
Figure 1.12: Investment Probability by Size, 1998



Notes: Truncated at 2500 MW. Sizes have been divided into 25 bins. Requires investment to be at least 20% of existing capacity. Investments are recovered from looking at year 2000 since 1999 is missing. Graph covers all provinces. Data comes from capacity panel in confidential physical power plant survey.

A major takeaway from this graph is that a firm's size does not appear to be initially a huge predictor of whether it will invest or not. This continues in 2003:

Figure 1.13: Investment Probability by Size, 2003



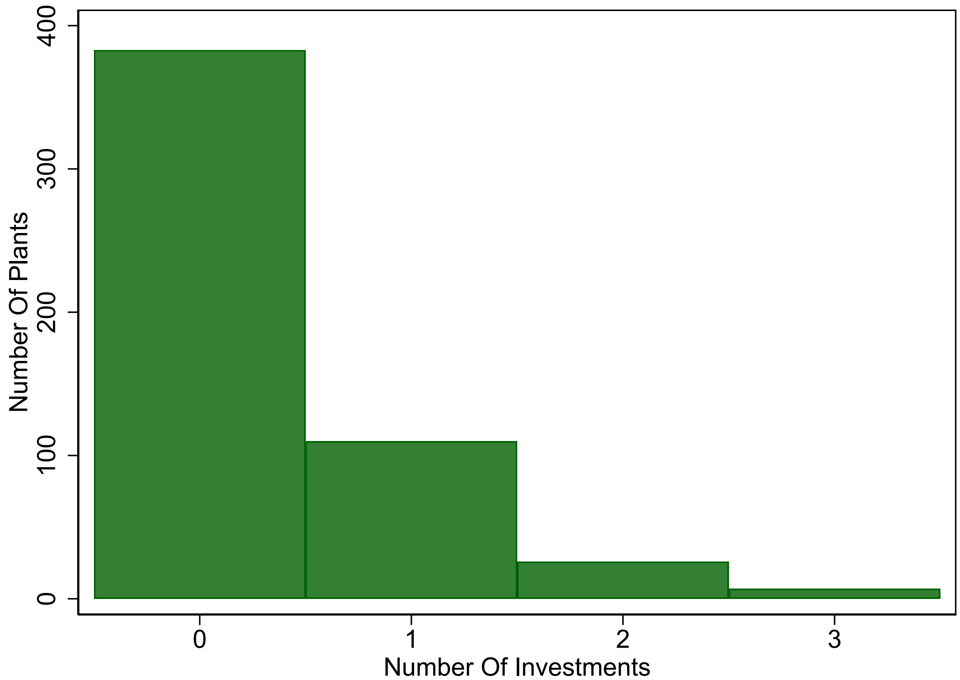
Notes: Truncated at 2500 MW. Sizes have been divided into 25 bins. Requires investment to be at least 20% of existing capacity. Graph covers all provinces. Data comes from capacity panel in confidential physical power plant survey.

Overall, investment rates are substantially lower across the board than in 1998. There are many more small investments, but otherwise there is still no discernible relationship between size and investment.

Under many traditional models of investment, returns to capacity are diminishing (see, ie, Ryan (2012)). A plausible expectation would be for plants to invest until they reach a target size such that the return has leveled off, and then stop investing. This would imply a negative relationship between size and investment probability, yet this graph suggests either no systematic relationship, or possibly a weakly positive one. There are several plausible explanations for this phenomenon, such as the fact that plants may be operating in a non-stationary environment, or that there is unobserved heterogeneity in investment costs.

It is straightforward to test for the presence of unobserved heterogeneity. A model-free fact that would support this hypothesis would be if there were plants with more

Figure 1.14: Distribution of Investment Episodes, 1998-2007



Notes: Requires investment to be at least 20% of existing capacity. 1999 and 2001 are missing. Graph covers all provinces. Data comes from capacity panel in confidential physical power plant survey.

investments than others: in this case, these potential "low cost" plants would be exploiting their status, while the vast majority of plants simply face costs too steep to make investment ever worth it.

At the plant level, roughly half of observations never invest. For plants that do invest, the majority of them only do it once. However, at least 50 investment (with a cutoff of at least 20% of existing capacity) episodes are done by plants who invest more than once. Some of this may be due to "time to build" that is longer than one year, or small amounts of measurement error since the investment panel has more than one underlying source.

To get a better understanding of what states are necessary to incorporate in a dynamic model of investment, it is important to investigate which variables influence a plant's investment decision. Since investment increases a plant's size, we would expect some kind of systematic relationship between size and investment. Given that investment is lumpy and rarely done more than twice by any plant, this suggests there is some kind of target size plants would like to reach and stay at for at least a few years.

Table 1.4: Regressions of Investment on Current Size and Cost

VARIABLES	Investment (MW)				Investment (Binary)	
Capacity (MW)	0.0180*	-0.544***	.025**	-.54***	-2.3 e-05	-6.0 e-04***
	(0.00949)	(0.0274)	(.01)	(.03)	(1.67 e-05)	(5.09 e-05)
Constant	10.88	310.4***	-6.7	296***	.107	.50***
	(83.57)	(19.53)	(84)	23	(.147)	(.04)
Marginal Cost			94.7**	77		
			(45.6)	(68)		
Fixed Effects	Year + Mkt	Year + Plant	Year + Mkt	Year + Plant	Year + Mkt	Year + Plant
Observations	1,498	1,498	1,498	1,498	1,498	1,498
Number of Plants	428	428	428	428	428	428
R-squared	0.050	0.275	.052	.276	.059	.131

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Investment variables are derived from capacity differences over time in confidential power plant survey. Binary investments are conditional on being at least 10% of existing capacity.

A regression of a plant's investment size on its current size, conditional on investing, suggests a **positive** relationship between size and investment. While this is not fully robust to a linear or binary specification, significant negative relationships with size only appear in the presence of plant-level fixed effects. Additionally, the fit improves significantly with the inclusion of individual fixed effects across specifications. Thus far the evidence is consistent with substantial unobserved heterogeneity. This effect is equally pronounced when marginal cost is included. Similarly, a positive relationship between marginal cost and investment is observed without the inclusion of fixed effects. In the fixed effects specification, marginal cost has no effect. This may be puzzling at first: one would expect more efficient plants to forecast more favorable returns, and thus invest more aggressively. However, the documented output misallocation in this market may neutralize this. More efficient plants may not actually receive more favorable profits, and a lack of correlation between marginal cost and investment is consistent with this story.

Plants are operating in a non-stationary environment, so there is not, ex ante, an obvious relationship to predict regarding size. However, all specifications include year and at least market-level fixed effects, so to the extent that payoffs vary systematically over time, that variation is controlled for.

1.3 Model

The reduced form results provide an inventory for what is needed for a structural model to back out the dynamics of misallocation in this market: A static model capable of isolating plant-year level misallocation "wedges" from unobserved cost shocks, and a dynamic model that accounts for a plant's capacity, physical marginal cost, expected wedges, province and year, and unobserved heterogeneity in investment costs. The allocation model, which estimates planner policy functions and thus plant-level payoff functions, serves as "step one" in the analysis. It is through this channel that I document the first estimates of output misallocation in this market, and measure the possible gains from efficient reallocation.

1.3.1 Overview and Notation

Power plants [4] i exist in each market m and are allocated production in each period $t = 1, ..., \infty$. Except where necessary I suppress the market index. Each market is governed by one planning authority who delegates production q_{it} for each plant, which is determined by a plant's cost cc_{it} and wedge μ_{it}. cc_{it} consists of a physical component HR_{it} based off of a plant's heat rate, and an input price w_{it} equal to the plant's cost of buying coal. Planners observe each plant's cost and capacity cap_{it} and then make a continuum of discrete choice decisions each period to determine each plant's q_{it}.

Electricity import and export exists between markets but is assumed to happen at rates taken to be exogenous by both planners and plants. Each market contains a composite fringe plant, 0, which is used for normalizations in the static discrete choice framework. Depending on the market, the fringe consists either of aggregates plants for whom pricing data is not available but production data is, and/or a representative "importer" firm whose qualities are imputed from surrounding markets and is assigned production equivalent to a province's imports.

Prices p_{it} are assigned exogenously by planners via an approximation of cost-of-

[4] For the purposes of the model, all decisions are made at the plant level, rather than the firm level. There are several reasons behind this: First, my understanding is that most privately owned plants are actually the sole plant owned by the operating firm in many cases. Second, for cases of common ownership, like large, state-owned firms, their plants tend to be dispersed across many markets. For example, in 2006, the smallest one of these firms had plants in at least 10 distinct markets, many of which are cut off from each other due to China's transmission infrastructure. Third, within a province these plants are likely far from each other, and the government generally has enough power over both prices and quantities that plant dispatch decisions would be assigned mostly separately. Taken together, these all mean that plant-level decisions with some possible shifters for state-owned firms should capture almost all of what is happening with regard to production and investment.

service regulation based on plant-level characteristics cc_{it}, cap_{it}, and μ_{it}. p_{it} and q_{it}, both of which are fully determined by the planner based on factors that are pre-determined each period, combine to form a plant's period payoff π_{it}. μ_{it} is allowed to depend directly on cc and cap in the same manner. μ and cc evolve according to interdependent AR(1) processes.

Plants make investment decisions in a single-agent, discrete time, infinite horizon framework. Investment is a plant's decision to expand its operating capacity cap_{it} or not. This is assumed to take 3 possible values, such that a plant can either not invest, make a "small" investment, or a "large" investment each period [5]. Small and large investments come with associated costs $\gamma_i(x)$, where x denotes the amount a plant invests, and i represents a plant's categorization into one of multiple possible discrete types of investment cost functions. Apart from investments, cap_{it} evolves deterministically. Investment decisions are made to maximize expected lifetime profits, which are determined by a plant's sequence of payoffs $\pi_{it,t=1}^{t=\infty}$ and Rust (1987)-style logit shocks $\varepsilon_{it,t=1}^{t=\infty}$ over their investment choices, as well as any costs they incur from investing.

Entry decisions are treated as exogenous, but plants who do enter are assumed to make an initial investment decision to determine their starting capacity as well. Apart from a different cost structure $\gamma_i^e(x)$ and a different information set prior to their first investment decision, these investments are made under identical conditions as incumbents. Exit is largely exogenously determined during the span of my sample and I abstract from it in the current iteration of the paper.

Plants receive static payoffs π_{it} over an infinite discrete time horizon. Aggregate

[5] Ideally the model would have a finer set of choices, but computational demands require the choice set be kept fairly small. 3 possible choices allows for two fundamentally different "kinds" of investment, between modest additions to small plants and major upgrades to major sites. Thus, while I continue to experiment with other sizes, this appears to be a reasonable compromise.

demand Q_{mt} is taken to be exogenous by each planner and plant, and both are assumed to have perfect foresight over its evolution. The same assumption holds for the sequence of entry and exit within each market. Plants approximate these variables as well as any possible policy changes by forecasting their payoffs under non-stationary conditions in the following way: for the first 10 years $t = 1, ..., 10$ plants have perfect foresight over aggregate conditions, represented by the inclusion of t as a state variable. Letting $t = 10 = T$ approximate a sort of "terminal period", plants then forecast their payoffs to be stationary with a value function that assumes payoffs are thereafter determined by the payoff function from period T.

1.3.2 Timing

The model can be summarized by the following sequence of intra-period steps:

1. Plants are allocated production via μ in period t and assigned prices based on their key observable variables (marginal cost, capacity), as well as by province and year.
2. Random investment shocks ε are realized, and plants make investment decisions x. Plants maximize their expected lifetime stream of profits, and forecast their future payoffs.
3. Entrant plants make initial investment decisions.
4. Profits in period t (including investment shocks and costs) are realized.
5. Investments realize and time iterates, with Q_{mt} as well as any other aggregate state variables evolving deterministically.
6. Period $t+1$ starts with new capacities, costs and values of μ.

Let $s_{it} = (\mu_{it}, cc_{it}, cap_{it}, t, m)$. To summarize (with a slight abuse of notation), a plant of a given cost type solves the following decision to determine its optimal investments:

$$max_{x \in X} \pi(s) + \varepsilon(x) - \gamma(x,s,k) + \beta EV(s' + x|s) = V(s) \tag{1.1}$$

Where $\pi(s)$ represents static period payoffs, $\gamma(x,s,k)$ are investment costs for discrete type k, $\varepsilon(x)$ is a choice-specific logit shock, and $\beta EV(s'+x|s) = V(s)$ are expected future profits.

1.3.3 Components of Per-Period Payoff Function

Regulator Quantity Policy Function

Each plant is ranked according to a cost index b_{ith} in the provincial regulator's dispatch process:

That is:

$$b_{ith} = \mu_{it} - cc_{it} + \varepsilon_{ith} \tag{1.2}$$

cc represents a plant's physical marginal cost. μ is a plant-level index that can persistently alter a plant's expected rank up or down, causing it to produce differently than if dispatch were conducted purely based on marginal cost (and shocks). h represents an infinitessimally small time interval over which the regulator makes a continuum of decisions to allocate production to each plant.

Under optimal dispatch, μ would be 0 for all plants and their bids would be purely based on cc, with the lowest cost plant producing first until its capacity is filled. μ is the vector of parameters of interest–it represents a plant's misallocation "wedge" much like in Hsieh and Klenow (2009). μ may represent persistent political connections, flawed infrastructure, corruption, or a reflection of an unobserved objective function the dispatcher has.

ε captures unobserved shocks that the regulator experiences. These may be due to unobserved costs or political edicts. Letting ε be a standard Type 1 Extreme Value shock yields the standard logit form:

$$P(b_{ith} > b_{jth} \forall j \neq i) = \frac{exp(\frac{\mu_{it} - cc_{it}}{\sigma})}{\sum_{jt} exp(\frac{\mu_{jt} - cc_{jt}}{\sigma})} \tag{1.3}$$

To identify μ for every plant, normalizing $E(b_{it}) = 0$ for one plant in each market is necessary. Unlike in the standard consumer discrete choice context, there is no obvious "outside option" to normalize like simply not buying a product. Since supply and demand have to be continuously and exactly equal in the electricity industry, there is never any observed "excess" production.

Instead, I take all of the production that is not matched to the financial census, aggregate these plants (which are on average smaller) into a "fringe", and treat their aggregate properties as one additional plant in each market. I also impute a cost and capacity for the decision to import electricity from neighboring provinces, and incorporate this into the properties of the additional plant if a province both has imported electricity and fringe plants. For provinces that only contain one I use that as the fringe plant. This way all province-year combinations have a feasible outside option. This fringe plant captures a market-level average cost that electricity can be produced at rather than any plant in question.

To account for a plant's hard capacity constraints, suppose additionally that if a plant wins a portion $\int_i$ of the continuum of auctions, they will be allocated an annual production of $cap_i * \int_i$. According to the logit structure of the unobserved cost shock, for market size Q, borrowing algebra from Berry (1994), this would result in the following observed production share:

$$\frac{q_{it}}{Q_t} = \frac{cap_{it} exp(\frac{\mu_{it} - cc_{it}}{\sigma})}{cap_{0t} + \sum_j cap_{jt} exp(\frac{\mu_{jt} - cc_{jt}}{\sigma})} \tag{1.4}$$

The key difference from this and the standard Berry (1994) logit model is that each observed production share has to be weighted by capacities. Thus, while each plant's probability of winning follows the standard logit formula, their observed production share has to take into account the full set of capacities within their market [6]. The fringe plant, 0, is still normalized to have a bid of 0, but their contribution is now weighted by their capacity.

This admits the following representation:

$$ln(\frac{q_{it}}{cap_{it}}) - ln(\frac{q_{0t}}{c_{0t}}) = \beta_0 + \mu_{it} - \beta_1 cc_{it} \tag{1.5}$$

Where, as before, the coefficient on the physical marginal cost represents the inverse of the variance of the unobserved cost shocks. This equation represents the regulator's policy function for allocating production in each market.

[6] This follows quickly algebraically: if $q_i = s_i * cap_i$ and s_i is determined by the logit probability process, then each plant's q is simply its capacity times its logit probabilty. Summing over each plant and imposing market size Q generates this result. It is also important to note that Q does not represent the exact size of the continuum of decisions like M does in the consumption context. Rather, the size of the continuum of choices is a function of the distribution of capacities in the market and Q, which the planner realizes to meet demand

Finally, to account for time invariant portions of a plant's wedge (such as a persistent political connection), μ_{it} will be separated into a fixed and random effect: $\mu_{it} = \mu_{fi} + \mu_{rit}$. Fixed effects may also capture unobserved heterogeneity in plant cost structure.

Constructing Fringe Plants

A key component of the planner's decision is the normalized fringe plant in each market. This representative firm has two possible components: within-market fringe firms, and/or an "importer" firm.

Within-Market Fringe The construction of the within-market fringe firm depends on additional production data outside of the "revenue" sample. For most markets, there are power plants for which I do not observe financial information and cannot explicitly include in the dynamic analysis. However, I observe both their capacities and their production levels, so they can be included in aggregate in each planner's decisions.

Thus, denoting q_i^{wf} as the production of each within-market fringe firm and cap_i^{wf} as their capacity, their utilization is calculated from $\sum_i q_i^{wf}$ and $\sum_i cap_i^{wf}$.

Import Fringe Production While many provinces in China are isolated with regards to their electricity production and consumption, some inter-provincial transmission does take place. According to CEIC data, Beijing in 2003 produced roughly 1900 MW of electricity, and imported another 3500, since the entire province is one large city with little room for power plants that would produce enough for its dense population.

Meanwhile, Inner Mongolia exported roughly 3500 MW that year. The two provinces have fairly similar populations, but Inner Mongolia is much more rural and rich with coal. While cross-province transmission infrastructure is generally weak in China, there are cases like these where it will be important to account for this.

There are severe data limitations on import and export data: I do not directly observe which plants import or export, where each province imports or exports to on average, or what the actual transmission network looks like. I impute these values using aggregate data from CEIC and a few key assumptions:

Assumption I1: The provincial dispatcher treats imported electricity as though it is coming from one composite "plant". When production is being allocated, the simulated "importer plant" has its own imputed costs, capacity, and μ.

Assumption I2: The composite importer plant inherits the properties of other provinces via inverse distance weighting. If all imports to a province are going to be treated as coming from one plant, it is necessary to take a stand on that plant's characteristics.

Given China's week transmission infrastructure, the vast majority of imported electricity likely comes from neighboring provinces, as this is the simplest option requiring the shortest length of power line. This is not the exclusive means by which electricity is transmitted, ever. "Two stage" transmission of electricity, where power travels through an intermediate province en route to its destination, does happen in China [7]. Thus, taking aggregate imports and apportioning them to other provinces as some kind of decreasing function of distance is an attractive approximation.

To accomplish this, I take every pairwise combination of distances between the capitals of each province, and take the inverse of its square. I then normalize each of these

[7] For example, West to East transmission is referenced in Peng (2017).

objects by the total sum, and this constitutes a province's "weight" in the importing province's total imports [8].

Put more simply, let d_{ij} be the distance between province i and province j. Without loss of generality, let i be the importing province. The import weight w_{ji} from province j in province i is thus:

$$w_{ji} = \frac{\frac{1}{d_{ij}^2}}{\sum_{j' \neq i} \frac{1}{d_{ij}^2}} \tag{1.6}$$

This approach works well for large parts of China. Take, for example, Chongqing which is a smaller province in the middle of China. The weighting imputes that 20% of imports would come from Guizhou, 7% from Hubei, 21% from Sichuan, 6% from Shaanxi, 5% from Hunan, and under 5% from all other provinces. It may well be that even more of Chongqing's electricity imports come from these nearby provinces, but in terms of figuring out the average cost and capacity reflected by these import numbers, they are clearly assigned a much larger weight than distant areas.

There are a few instances that may require more specific assumptions. For example, this approach leads one to assume much of Beijing's electricity is imported from Tianjin. This may be true, but it is unlikely given that these are both large cities. More extreme outcomes, like electricity being transmitted across the country from Heilongjiang to Yunnan, are unlikely, as this imputation assigns only a .4% weight for that combination.

[8] While these weights vary over time and take into account changes in each province's capacity and cost makeup, this approach is not fully consistent with the accounting nature of the underlying data yet. To do that, I need to restrict the weights to only come from provinces that exported in a given year, which makes the problem somewhat more complicated. For now, I stick with the version of the imputation for simplicity's sake.

With w_{ij} in place for all possible combinations of provinces, it is possible to compute weighted averages of properties from the exporting provinces. For capacity, I take each province and add up the total capacity across plants (in each year), $C_i = \sum_f c_{if}$ where i indexes provinces and f indexes plants. Then:

$$c_{imp,i} = \sum_j w_{ji} C_j \tag{1.7}$$

Costs are a similar calculation, but with mean linear marginal costs rather than total ones. Each importing province has a production amount, capacity, and marginal cost associated with its "importing plant", and they can be plugged into the model with these properties as though they were a standard power plant.

Beijing imports a high percentage of its electricity every year, as mentioned earlier. According to this analysis, the imputed importer marginal cost for Beijing in 2003 is .11, with an imputed aggregate available capacity of 13,325 MW. By comparison, the (unweighted) mean MC in Beijing for that year .08, and the total available capacity is only 2480 MW. This imputation thus supports the notion that Beijing has a limited amount of internal capacity available, and marginal costs are not so much higher in neighboring provinces that importing would be prohibitively expensive.

Fringe Utilization Total fringe utilization is thus the following sum:

$$\frac{q_{m0}}{c_{m0}} = \frac{q_m^{wf} + q_{imp,m}}{c_m^{wf} + c_{imp,m}} \tag{1.8}$$

Exports

Exporting provinces require a different set of assumptions given that I do not observe which plants export nor the destination of exported electricity.

Assumption E1: The provincial authority only cares about their trade balance, not import and export separately. This is to say that in reality most provinces do some amount of importing and some amount of exporting, likely because transmission networks vary in quality within and across provinces. For the purposes of the structural model, this means that province-year combinations will get categorized as either importers or exporters.

Assumption E2: Only the plants observed in the financial census data export. This assumption implicitly says that since there is a revenue cutoff to be included in the financial census, only these plants would be economically important enough to the province to be connected to the parts of the transmission network that are capable of crossing provinces.

Assumption E3: Plants that export do so in proportion to their observed production share. Given Assumption E2, which decides who exports, Assumption E3 says how exports are distributed. This is effectively a capacity reduction for every plant in the province. The planner has a certain amount of demand to meet, but unobserved constraints (be it federal policy or infrastructure deficiencies) force them to export power. Given a market-level export number E, each plant's effective capacity is reduced by $\frac{q}{Q}*E$.

Prices

For output prices, regulators care about plant costs and may engage in some amount of negotiation, but prices are more or less exogenously and idiosyncratically determined based on plant observable characteristics otherwise.

Thus, prices as modeled as such:

$$p_i = \alpha_0 + \alpha_1 cc_i + \alpha_2 ln(cap_i) + \varepsilon_{pi} \tag{1.9}$$

In practice, prices and quantity are combined to form a revenue function, and an approximation of this function is estimated to simplify dynamic estimation. See Appendix for details.

1.3.4 Evolution of State Variables

Thus far the model relies on five state variables: (cap, μ, cc, m, t). m, being a firm's market, never changes and is the same for every firm within a province. The remainder of the section outlines how the other states transition:

Capacity: Capacity depends directly on a firm's investment, and otherwise evolves deterministically. Taking a firm's current capacity to be cap_{it} and their investment to be x_{it}, the capacity transition process is $cap_{it+1} = cap_{it} + x_{it}$.

t: The transition process for the aggregate state variable, t, does not have a delineated transition function. Rather, its values are estimated nonparametrically in each year's payoff functions before the dynamics are estimated, and firms are assumed to have perfect foresight over its evolution. Formally, there is a sequence of known payoff functions $\pi_{1998}, \ldots, \pi_{2007}$, which receive the other four state variables as inputs each year.

Continuous States: Marginal costs are assumed to be persistent and responsive to both mechanical and financial input cost shifts from changes in capacity:

$$cc_{it} = \zeta_0 + \zeta_1 cc_{it-1} + \zeta_2 cap_{it} + \varepsilon_{cit} \tag{1.10}$$

Via my timing assumptions, costs and capacity are determined prior to μ. This is because the planner observes the operating capabilities of each plant, and is assumed to delegate production based off of both this and other, unobserved objectives. Thus, cc and cap can influence μ contemporaneously:

$$\mu_{it} = \tau_0 + \tau_1 \mu_{it-1} + \tau_2 cap_{it} + \tau_3 cc_{it} + \varepsilon_{\mu it} \tag{1.11}$$

The two errors terms may be correlated in estimation.

1.3.5 Value Functions and Equilibrium Concept

Equilibrium: A plant's investment decision depends on its values for μ, marginal cost, capacity, price, and relevant market-level states. The share/revenue equations imply the market states would be Q_{mt}, aggregate production, and s_{0mt}, fringe utilization, which implicitly depends on the behavior of every plant in the market, and any market or year level fixed effects in any regulator policy functions or costs.

Modeling and simulating this as a standard Markov Perfect Equilbirum model is infeasible, as there can be dozens of firms in a given market. Nonstationary Oblivious Equilibrium, from Weintraub et al. (2017) or the moment-based Markov equilibrium from Ifrach and Weintraub (2017) would be ideal candidates for approximate dynamic oligopoly methods. But, both would require taking an explicit stand on the aggregate states that influence both firm and planner decisions. This is particularly difficult in my setting: output and input prices, both of which would be necessary to include, are missing for large portions of my sample. Thus, accurately measuring average output price, for example, is next to impossible.

However, with the presence of t as a state, it is fairly straightfoward to motivate a plant's investment decision as a single agent model. A plant's static payoffs only depend indirectly on other agents, through whatever rule determines their μ. While a full specification of this rule and a full tracking of every other plant's μ would be accurate, from a plant's perspective two things are true: First, the allocation of μ is likely to depend heavily on macro shocks that the plant has little influence over. Second, μ is likely persistent and also dependent on a plant's observable characteristics. This suggests that if a plant tracks its own observables (marginal cost, capacity) and its value of μ within a model that allows for macro shocks, it is an adequate summary of all the payoff-relevant information that a plant is forecasting over.

Thus, I do not model plants as explicitly acting strategically, but rather I use single agent methods from Rust and Phelan (1997) and Conlon (2012) to build a model that captures the important macro variation in China, gaining the benefit of not having to specify a specific set of observables for aggregate market-level states. My equilibrium concept comes with an assumption on how macro shocks vary, which requires that plants forecast that the market becomes stationary at the end of the sample.

Finally, as suggested by the reduced form evidence, there is likely considerable unobserved heterogeneity in investment costs. I resolve this via a mixture model with two discrete types and the EM algorithm (see, ie, Scott (2013) or Arcidiacono and Miller (2013)).

Value Functions: Each period, plants make the following investment decision:

$$max_{x \in X} \pi_t(\mu, c, cc, m) + \varepsilon(x) - \gamma(x, cap) + \beta EV_{t+1}(\mu', c+x, cc', m|\mu, c, cc) = V_t(\mu, c, cc, m) \tag{1.12}$$

Where V represents the plant's optimal value function and EV is the expected next-period value of that function. $\gamma(x, cap)$ represents a capacity adjustment cost function. Note the t subscripts indicate a potentially non-stationary environment. $\varepsilon(x)$ is the standard logit shock seen in Rust (1987). As in that paper, this should be interpreted as a state variable that the plant observes and the econometrician does not, which affects the cost of investment stochastically.

This permits a standard representation of a plant's expected profits:

$$EV_t(s) = log(\sum_x exp(\pi_t(s) + \beta EV_{t+1}(s+x))) \tag{1.13}$$

Where, in a slight abuse of notation, s represents a plant's full vector of states, and x represents their investment decision.

Initial Investments

Given the large amount of entry into the market over this time period, I also incorporate the investment decision of entrants. I assume for now that entry is exogenous, and plants merely choose their initial level of investment.

Plants are assumed to forecast based on the province-year level mean profits, as well as any cost savings due to choosing a larger capacity. The entrant's problem thus becomes:

$$V^e_{m,t} = \max_x -\gamma^e(x) - \varepsilon(x) + \beta EV_{m,t+1}(cap'(x), \mu'(x), cc'(x)|x)) \tag{1.14}$$

I assume that entrant investment costs are purely linear, and scaled from incumbent ones by a parameter ρ_e.

1.4 Methods Used To Solve and Estimate the Model

1.4.1 Allocation Model

Production Allocation: Quantity assignment can be estimated via IV regression with instruments for physical cost. Two instruments that work well together are lagged physical cost, and the most recent amount invested by the largest plants in a plant's market. The reasoning behind the first instrument is straightforward: since cost is somewhat persistent yesterday's cost should be correlated with today's cost. But, contemporaneous shocks to *cc* that would be correlated with the residual of this regression should not be correlated. This will likely require a serial correlation SE correction.

The investment instrument captures aggregate cost shifts: the large plants in a market likely invested because they received favorable cost draws or misallocation wedges, and this would only be correlated with a plant's physical cost through a common aggregate shock. This common aggregate shock should not affect relative allocations the next period if all plants experience it, and thus this instrument should be correlated with cc and not μ.

Prices: Price coefficients are estimated using linear IVs and the same instruments as in the quantity allocation model.

1.4.2 Revenue Function Approximation

To reduce the state space for dynamic estimation, I first generate a 25 x 25 x 25 grid of the three "micro" states (cap, cc, μ). I also add each observed quantity of $ln(s_0)$ and Q for each year and province (which varies). In total this creates 2,937,500 points.

I then perform OLS of the log of the fitted revenues on log capacity, cc, μ, and a series of year and province dummies. This results in an R^2 of .97. To eliminate η and have prices be fully determined by the states, I draw a simulated price value for each evaluated function point as well.

This fit suggests that a non-stationary dynamic problem, where value functions are allowed to vary by province and plants perfectly forecast aggregate shocks, will capture almost all of the necessary variation that a strategic model (or one that depends on aggregate states) would.

1.4.3 Continuous State Variable Evolution Processes

Both transition processes (cc and μ) are assumed to be continuous, and largely dependent on variables that are pre-determined. While capacity and lagged values are fully determined at the start of each period, it contemporaneous shocks may determine both μ and cc for some plants.

Thus, a variety of methods are used to estimate these processes: OLS for a baseline specification, IV regressions to examine endogeneity concerns in the wedge regression, and, finally, a Seemingly Unrelated Regressions specification to determine whether it is important to account for possible correlations between the error terms across the two processes.

1.4.4 Dynamic Solution Concept and Estimation

The dynamic model is solved using collocation methods along with value function iteration in the terminal period and backwards induction in all prior periods, and collocation methods to solve the value function for each plant type. Given the solved value functions for each type, the function values are fed into a Maximum Likelihood outer loop routine. The EM algorithm is used to determine the distribution of plant types.

Inner Loop

This can be thought of as incorporating nonstationary nested fixed point methods like Rust and Phelan (1997) into a continuous state interpolation framework. It is important to note that this is not a discretized grid like in the classic Rust paper, but rather a series of nodes at which I fit polynomials that I then interpolate between as in Fackler and

Miranda (2002).

In the second to last period of the sample, plants assume that there is a stationary environment where value functions are equivalent to that in T forever (in this case 2007). One can exploit this structure to develop an "inner loop" similar to the classic Nested Fixed Point algorithm, but with an added series of backwards induction problems that firms use given their terminal period value function.

The basic goal of the inner loop is to use the collocation method to find a series of basis functions that satisfy the fixed point representation of the Rust (1987) model above. This can be represented in the following way:

$$\sum_{rtp} \phi_{rtp}(cap_i, \mu_i, cc_i) = log(\sum_{x} exp(\sum_{r} E_{\mu,cc}(\pi_{tp}(cap) + \beta \phi_{rtp}(cap + x, \mu_i', cc_i')))) \tag{1.15}$$

The tp subscripts represent that value functions are allowed to vary by year and province, while r represents each grid point. In this specification, there are $RxTxP$ basis functions ϕ to be solved for. In some specifications, I include a finite mixture model in the investment cost function (γ) which doubles the number of necessary functions to solve, which can be denoted as ϕ^1 and ϕ^2.

For a solved V_T, we can do the following backward induction:

$$EV_{T-1}(s) = log(\sum_{x} exp(\pi_{T-1}(s) + \beta \sum_{r} \phi_{rT}(s + x))) \tag{1.16}$$

Where s represents the 3 individual-level states.

Rather than doing statistical estimation inside the inner loop, this equation is set up to be just-identified. That is, there are R basis functions and R node points, and I require for this equation to hold exactly true at these points. Given a choice for a number of points R, I populate the grid using Chebyshev polynomials.

Given a grid, I choose a set of initial guesses for ϕ and use function iteration to solve for the fixed point in the terminal period. After this, I use backwards induction to separately solve each year in the sample. All markets and types are done simultaneously using parallel computing methods.

Choosing R The grid size was chosen adaptively: starting from a small number of nodes, I increase the value along each dimension until the estimates stabilized. This results in an 18 x 10 x 10 (or 1800 nodes) grid for capacity x μ x cc.

Restricting the Range Doing Chebyshev interpolation requires pre-specified bounded intervals to set up where the grid points are. Given that I cut off my sample at 50MW, I use this as the lower bound for capacity, and 4800 for the upper bound, which roughly corresponds to slightly more than the maximum observed capacity plus the largest investment in the sample, such that plants could forecast investing to this amount realistically.

Discretizing the Investment Space While cc and μ are treated as continuous and allowed to evolve via AR(1) processes, it is still necessary to discretize the investment decision. Binning investments into two sizes of "medium" (400 MW) and "large" (1000 MW) fits fairly well for now. Investments under 200 MW are binned to 0 in this specification.

The combination of these methods permits a value function for every type, market, year, and combination of individual state variables.

Outer Loop

The outer loop requires minimizing the likelihood function implied by the continuous Rust model I have set up. This is where the set of capacities and investments directly observed in the data is heavily exploited. Given an EV function as in Rust, this results in the following likelihood for a given choice:

In my case, the EV function is only known to be exact at a finite number of points. But, given the basis functions, we have an approximate answer from the solution in the inner loop that can be plugged in for the observed choices:

$$P_{tp}(a = x|s) = \frac{exp(\pi_{tp}(s) - \gamma(a) + \beta \sum_r \phi_{rtp}(s+a))}{\sum_x exp((\pi(s) - \gamma(a) + \beta \sum_r \phi_{rtp}(s+x)))} \tag{1.17}$$

Given this representation for a single observation's likelihood, it is straightforward to take the product for each power plant and maximize the log sum of these probabilities.

Incorporating Entrants The specification of an entrant's initial investment problem admits a probability of investment conditional on a plant's type, which is incorporated into its likelihood contribution along with the series of additional investments the plant makes. I discretize the entrant investment grid at 100, 300, 600 and 1200 MW based on the empirical distribution of initial investment decisions, and impose that $\sigma_e = .75 * \sigma_i$ to account for the higher number of discrete choices.

Likelihood Functions: A plant's likelihood contribution is expressed by $P_{tp}(a = x|s)$ above. Since there are discrete types, it is necessary to consider a plant's likelihood conditional on their type k. That is, $P_{tp}(a = x|s,k)$. The structure of their likelihood contribution is still the same, but different conditional likelihoods will occur as different dynamic parameter values are considered for each type.

For a plant that is observed the entire duration of the sample, I assume their initial capacity is exogenous. Their conditional likelihoods are simply:

$$L_i(z,\theta|k) = \prod_t P_{tp}(a_{it} = x_{it}|s_{it},k,\theta) \tag{1.18}$$

For plants that enter during the sample, they also have a likelihood contribution related to their starting capacity. This takes an identical form to the contribution for each incumbent investment decision, but depends on different components of θ. Additionally, no payoffs are incurred the period the capacity decision is made, and as above plants assume they are entering at province-year specific mean levels otherwise. Denoting the probabilty of this decision as $P^e_{tp}(a = x|l)$, these conditional likelihood contributions can be summarized as:

$$L_i(z,\theta|k) = P^e_{1p}(a_{ik} = x_{ik}|l) \prod_{t>1} P_{tp}(a_{it} = x_{it}|s_{it},k,\theta) \tag{1.19}$$

Where $t = 1$ represents the plant's first observed period. It is still necessary to convert the type-specific likelihoods to an unconditional likelihood function. Let $\hat{p}_k$ be an estimate of the unconditional probability of being type k. Using Bayes' theorem, the likelihood of being a type (in this case, type 1 of 2) conditioned on the observed data is:

$$P(k=1|z,\theta,p) = \frac{\hat{p}_1 L_i(z,\theta|1)}{\hat{p}_1 L_i(z,\theta|k) + \hat{p}_2 L_i(z,\theta|2)} = \hat{P}_1 \tag{1.20}$$

An analogous equation holds for type 2. The unconditional likelihood is thus:

$$l_i(z,\theta) = ln \sum_k \hat{P}_k L_i(z,\theta|k) \tag{1.21}$$

For a given set of probability estimates, $\hat{\theta} = \arg\max_\theta \sum_i l_i(z,\theta)$. To search over θ I use the BFGS algorithm (stock "fminunc" in MATLAB, which is a pseudo-Newton method very similar to the BHHH algorithm generally used in applications of Rust (1987)).

Given a new value of $\hat{\theta}$ and P, new guesses for p can be calculated, and one can iterate in this manner (the EM algorithm) back and forth until the estimates for all 3 objects converge.

Dynamic Estimation Summary

To summarize the combination of the 3 components:

1. Choose a vector of parameters (6 total, 3 for each type) and unconditional probabilities p_1 and p_2 to start.
2. Solve each Rust model, the series of backward induction decisions, and entry decisions to get the value function for each type and market.
3. Given the value functions, solve for each conditional likelihood function for each type and market.

4. Calculate conditional type probabilities and a conditional likelihood function, given the unconditional likelihoods.

5. Find a new vector of parameters that minimizes the unconditional likelihood function.

6. Solve for the new unconditional probabilities implied by the calculated conditional probabilities at the minimum.

7. Iterate until desired convergence level.

1.5 Empirical Results

1.5.1 Allocation Model

Quantity Allocation: Current results utilize lagged cost and total investment for the five largest firms in a plant's market, as described above:

Table 1.5: Allocation Model Estimates of $1/\sigma$

	(1)	(2)
Estimate	1.61	.235
SE	(.594)	(.094)
Individual FEs	Yes	No
Year FEs	Yes	Yes
Province FEs	No	Yes
First Stage F	18.01	567
J Stat p-value	.30	.07
N	1501	1501

Notes: σ is variance of unobserved cost shocks, or the coefficient on marginal cost. $1/\sigma--> 0$ would imply perfect merit-order dispatch. Dependent variable is log utilization minus log fringe utilization.

The results imply that individual fixed effects dramatically affect the results. This is likely for two reasons: one, there is genuine unobserved heterogeneity in the type of plants that are operating. For example, even controlling for size, there will be "peak" and "baseline" plants that are meant to serve different parts of the demand curve. Second, there is likely a time-invariant component of misallocation. If the wedges include, say, political connections because of who a plant's manager is, then these would not be expected to change over the timeframe of the sample.

An estimate for σ of around .62 suggests that the unobserved cost shocks that realize through the year are extremely important. Given an average marginal cost of around .19, intra-annual demand considerations are clearly a large factor in determining which plants get allocated production.

Prices: Instrumenting for costs using the same instruments as above and including year and fixed effects on prices nets the following result:

Table 1.6: Pricing Regression Estimates

VARIABLES	(1) α_0	α_1	α_2
Estimate	.22	1.03	-.03
SE	(.07)	(.15)	(.008)

Notes: α_0 is intercept, α_1 is coefficient on marginal cost, α_2 is coefficient on log capacity.
Data sources include both NBS census and confidential survey.
First stage F stat is 19, J stat p-value is .11, N = 1501.

Pricing regression estimates suggest that prices hew very closely to marginal costs: α_0 represents a combination of what is likely a close to constant markup, akin to a return given in cost-of-service regulation, while $\alpha_1 = 1.03$ suggests that prices and costs vary almost perfectly linearly. A significant coefficient for α_2 means that large plants are actually penalized in the prices they are given on top of having lower costs, which forms an additional potential "wedge" from the planning process.

1.5.2 State Evolution Processes

The evolution processes of the continuous states are estimated using a variety of techniques: standard OLS, IV regression, and SUR estimation:

Table 1.7: State Transition Processes

Parameter	Separate OLS	Separate OLS	Separate OLS	Separate OLS+IV	SUR
Cost					
ζ_0 (Intercept)	.06 (.005)***	.19 (.01)***	.19 (.02)***	.19 (.01)***	.12 (.01)***
ζ_1 (Lag)	.78 (.02)***	.71 (.02)***	.67 (.02)***	.71 (.02)***	.71 (.02)***
ζ_2 (Log Cap)		-.02 (.002)***	-.03 (.002)***	-.02 (.002)***	-.02 (.002)***
Wedge					
τ_0 (Intercept)	.01 (.01)	-.003 (.06)	-.03 (.05)	-.07 (.08)	-.06 (.05)
τ_1 (Lag)	.26 (.02)***	.19 (.02)***	.17 (.02)***	.17 (.02)***	.18 (.02)***
τ_2 (Log Cap)		-.02 (.01)**	-.016 (.01)**	-.02 (.01)**	-.02 (.007)***
τ_3 (Cost)		.89 (.07)***	.97 (.07)***	.96 (.14)***	.92 (.06)***
Year FEs	No	No	Yes	No	No
First Stage F				152	
Error Correlation					-.01 (BP Test p = .58)
N	1498	1498	1498	1498	1498

Notes: Section headers refer to dependent variables. Standard errors in parentheses. * = significant at 10%, ** = significant at 5%, *** = significant at 1%. Costs are instrumented with lags in the IV specification for wedges, while the cost regression is never instrumented due to timing assumptions. Dependent variables are fuel costs and estimated wedges from allocation model.

The coefficients are not particularly sensitive to the choice of specification once capacity is included in both regressions. Year fixed effects attenuate the persistence of marginal costs, but otherwise the AR(1) coefficient is consistent across regressions. μ is less persistent than may be expected given that it can represent lasting political considerations, and in future work I will present results with fixed and varying components of μ evolving separately.

A couple of conclusions are clear no matter the choice of model: marginal costs are highly persistent, and decrease with changes in capacity. Similarly, larger capacity and lower costs are both associated with lower allocated production. The positive effect of marginal cost is particularly strong, suggesting inefficient plants are being aggressively propped up.

A Breusch-Pagan test rejects the correlation of error terms across the two regressions, so my preferred specification is for the two separate OLS regressions, as there is little change when using the more complex IV and FE specifications.

1.5.3 Investment Cost Parameters

Following Cooper and Haltiwanger (2006), investment costs are parameterized as follows:

$$\gamma(cap, inv) = \gamma_x cap + \gamma_q inv^2/cap + \gamma_l inv + \varepsilon(x) \tag{1.22}$$

γ_x is allowed to vary by discrete type, which parameterizes the unobserved heterogeneity in investment costs. It is also necessary to estimate the variance of the logit shock σ_l. Current estimates are for a special case of the model: plants only consider μ as a scalar and do not differentiate between their fixed and random component. Additionally, these results do not incorporate the import and export models.

Table 1.8: Parameters + Costs for a 500 MW Plant Investing 400 MW

VARIABLES	Baseline	Mix	Mix + Entrant
γ_{l1}	\$230 Million	\$650 Million	\$550 Million
γ_{l2}		\$94 Million	\$104 Million
γ_x	-\$7.6 Million	\$7.5 Million	\$11 Million
γ_q	\$370,000	\$167,000	\$630,000
$\hat{p}_1$		.67	.55
$\hat{p}_2$		.33	.45
σ_l	\$51 Million	\$34 Million	\$34 Million
ρ_e			.8
Observations	1,172	1,172	1,172

Notes: γ_l denotes linear investment cost, γ_x denotes fixed cost, γ_q denotes quadratic costs.
Types differ only in linear investment costs ($l1$ and $l2$)
$\hat{p}$ denote probabilities of being each type in mixture models.
ρ_e represents cost-scaling term for entrants.

We can see that the introduction of the mixture model creates a large divergence in the linear term: "high cost" types essentially never invest, while low cost types have extremely cheap investment costs. One issue the mixture model faces without incorporating the initial investment decision is it may tend toward a corner solution, where all non-investing plants are grouped as one type, and all plants that ever invest as another. Thus, the $650 million figure for γ_{l1} may not be identified. The entrant specification brings the numbers to something more reasonable.

The IEA estimates that capital costs are generally in the low thousands per kW (or low millions per MW, and thus hundreds of millions per 100 MW). Depending on the specification, these estimates fall comfortably within or fairly close to these broad industry estimates.

Goodness of Fit

Table 1.9: Observed vs. Predicted Mean Investment Probabilities

Type of Investment	Observed	Baseline	Mix	Mix + Entrant
Invest 1000 MW Bin	2.3%	1.0%	3.1%	2.0%
Invest 400 MW Bin	6.4%	13.1%	8.5%	9.4%
Invest 0 MW Bin	91.2%	86.0%	88.4%	88.5%

Notes: Probabilities are weighted averages across both types, and pooled across years and provinces.

The baseline (non-mixture) model severely underpredicts large investment, which makes it a poor candidate for counterfactual simulations. Given its combination of more realistic parameter estimates, overall good fit, and richer set of possible predictions, the mixture model with initial investment decisions included is the best overall specification. All further results and counterfactuals are based off of these values.

Table 1.10: Observed vs. Predicted Mean Entrant Investment Probabilities

Type of Investment	Observed	Preferred Specification
Enter 1200 MW Bin	19.7%	12.4%
Enter 600 MW Bin	26%	20.1%
Enter 300 MW Bin	16.2%	18.6%
Enter 100 MW Bin	38.2%	48.1%

Notes: Results are from assuming entry decision is exogenous, but initial investment decision is not. Distinct from incumbent investment decisions in earlier tables. Probabilities are weighted averages across types, and pooled across all years and provinces.

The entry model is very coarse, but still fits entrant investment frequencies fairly well. Given its tendency to overpredict small investment and underpredict large investment, a specification will improve the fit in future work.

Table 1.11: Probability of Being Type 1 or 2 Based On Investment Behavior

Type of Investment	P(Type 1)	P(Type 2)
Ever Invest 1000 MW	0	1
Ever Invest 400 MW	.0003	.9997
Never Invest	.56	.44

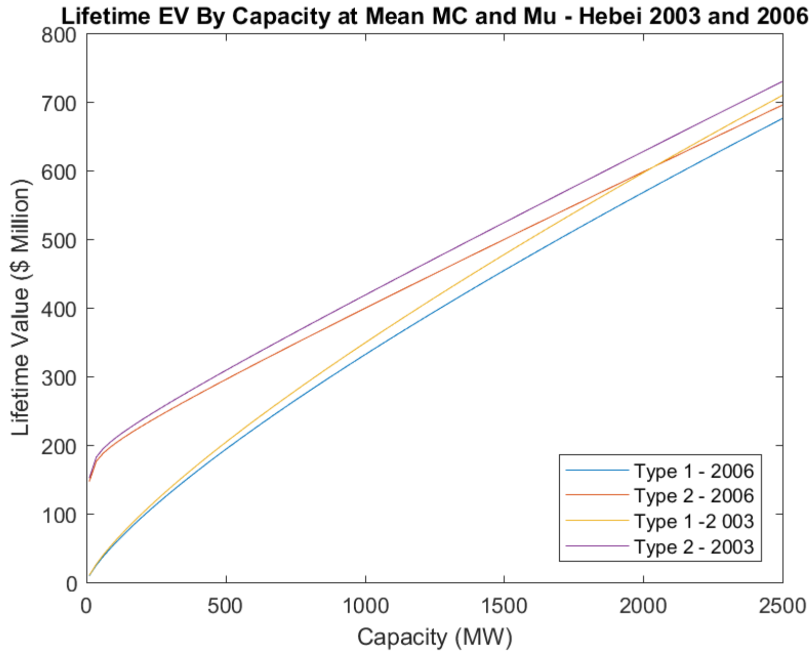
Notes: Type 1 is "high" cost, type 2 is "low" cost. Types are distinguished by their linear investment costs.

The mixture and entrant model still groups essentially all plants that invest deterministically into the low-cost type. But, non-investing plants will probabilistically have a fairly high chance of investing in counterfactual simulations as their chance of being low-cost is on average close to 50-50.

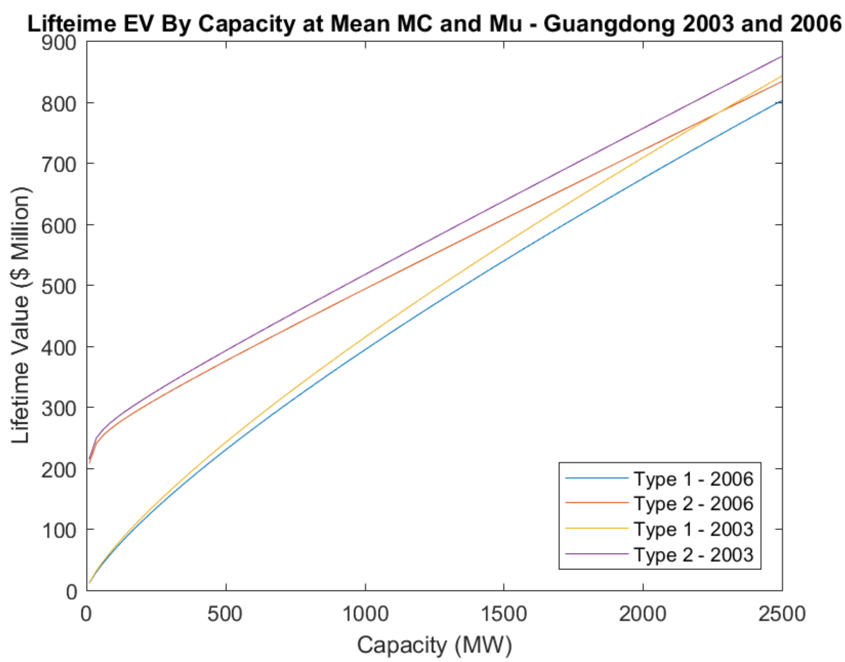
1.5.4 Value Function Estimates

Baseline Value Functions

Below are the estimated value functions by size at mean cost and μ values for two provinces: Hebei in the northeast, and Guangdong in the south:



Notes: Estimates reflect value function from dynamic model. Values are capped at 2500 MW as plants larger than this are not observed in this province/year combination. Values are for Hebei only. Mean μ is -.05, mean mc is .18 (000 RMB/MWh).



Notes: Estimates reflect value function from dynamic model. Values are capped at 2500 MW as plants larger than this are not observed in this province/year combination. Values are for Guangdong only. Mean μ is -.05, mean mc is .18 (000 RMB/MWh).

Two consistent takeaways emerge across provinces: the return to size is much larger at smaller capacities, and low cost types have much higher lifetime values than high cost ones. Provinces differ moderately in their slopes and intercepts, suggesting that solving the plant's problem separately for each market captured necessary heterogeneity.

Abstracting from provincial heterogeneity, below are some national average value functions for 2006:

Table 1.12: Predicted EVs - National Average (mean μ and cc) ($ Million)

Initial Size	Type 1	Type 2
100 MW	53.3	189.8
500 MW	185	280
800 MW	267	342
1500 MW	436	478

Notes: Results are averaged across provinces. Mean μ is -.05, and mean mc is .18 (000 RMB/MWh). Results are based on 2006 value function.

To get a sense of whether investment should be sensitive to either cost or misallocation, it is useful to see how values vary over these variables:

Table 1.13: Lifetime Return by Cost ($ Million) (500 MW, 06)

Initial Size	Type 1	Type 2
.2 (000 RMB/mwh)	184	279
.18 (000 RMB/mwh)	185	281
.16 (000 RMB/mwh)	187	283
.14 (000 RMB/mwh)	188	284

Notes: Mean μ is -.05, mean mc is .18 (000 RMB/MWh). Types are differentiated by linear investment cost terms, where type 1 is "high" cost.

The return to cost seems to be fairly low–but this may be due to misallocation getting in the way and preventing plants from investing as they otherwise would.

Returns to μ

The returns to an initial, one-time shock in μ are straightforward to calculate:

$$\%return = \sum_{m} \frac{EV_{tm}(\mu + SD_{mu}, cap, mc') - EV_{tm}(\mu, cap, mc)}{EV_{tm}(\mu, c + x, mc)} \tag{1.23}$$

Where cap, μ, and mc and t must be chosen beforehand. Varying these to commonly observed or modal values can give us an understanding of how μ differs in importance to different kinds of plants. Below are both absolute and percentage returns to a one-time, one SD increase in μ.

Table 1.14: Return on a 1 SD Initial Increase in μ (500 MW, 06, \$Million)

Initial Size	Type 1	Type 2
100 MW	3.64	4.65
500 MW	11.5	12.2
800 MW	16.2	16.6
1500 MW	25.4	25.7

Notes: Results are averaged across provinces. Mean μ is -.05, and mean mc is .18 (000 RMB/MWh). Results are based on 2006 value function. 1 SD of μ is equal to .4.

Table 1.15: Return (Percentage) on a 1 SD Initial Increase in μ (mean cc, 06)

Initial Size	Type 1	Type 2
100 MW	6.8%	2.5%
500 MW	6.2%	4.3%
800 MW	6.1%	4.9%
1500 MW	5.8%	5.4%

Notes: Mean μ is -.05, mean mc is .18 (000 RMB/MWh). Types are differentiated by linear investment cost terms, where type 1 is "high" cost. Results are averaged over provinces.

μ appears to matter more than cost in a plant's value function. A one-time (rather than persistent) change by 1 standard deviation of μ, averaged over provinces, changes a plant's value by 6% for high-cost types, and 3-5% for low-cost types. This may be purely due to higher expected static profits, but it likely also comes from altered investment behavior, which can be partially decomposed by looking at how a plant's investment policy function changes.

These can be calculated in the following way: A change in μ implies a change in EV, and from the dynamic model we can calculate investment choice probabilities from any given value of EV. Thus:

$$P(invest|s') = 1/m\sum_m P_{tmp}(a=x|s') = 1/m\sum_m \frac{exp(\pi_{tmp}(s'+SD_\mu) - \gamma(a) + \beta\sum_r \phi_{rtmp}(s'+a))}{\sum_x exp((\pi(s') - \gamma(a) + \beta\sum_r \phi_{rtmp}(s'+x)))} \tag{1.24}$$

s' can represent any change to a plant's states, but in this case it represents the addition of a 1 standard deviation shock to μ. The m subscripts denote that these values are averaged over provinces, since value function and choice probabilities vary by market.

Table 1.16: Investment Policy Functions in Response to a One-Time μ Shock, (500 MW, 2006)

Type of Investment	Baseline	-1 SD	-2 SD
Invest 1000 MW Bin	2.2%	2.0%	1.9%
Invest 400 MW Bin	6.1%	5.8%	5.6%
Invest 0 MW Bin	91.8%	92.1%	92.5%

Notes: Mean μ is -.05, mean mc is .18 (000 RMB/MWh). Results are averaged over provinces. One SD in μ is .4.

While the baseline probabilities are fairly small, even a one-time change in μ causes a plant to make a large or medium investment 5% (not percentage points) less frequently.

Compare this to a plant that is myopic about its stream of μ (and faces no uncertainty over it):

Table 1.17: Investment Policy Functions in Response a One-Time μ Shock, Myopic plant (500 MW, 2006)

Type of Investment	Baseline	-1 SD	-2 SD	+1 SD	+2 SD
Invest 1000 MW Bin	.14%	.13%	.12%	.15%	.16%
Invest 400 MW Bin	1.3%	1.2%	1.2%	1.3%	1.4%
Invest 0 MW Bin	98.6%	98.7%	98.7%	98.5%	98.4%

Notes: Myopic plants only react to present value of μ, and do not incorporate it into future value stream. Mean μ is -.05, mean mc is .18 (000 RMB/MWh). Results are averaged over provinces. One SD in μ is .4.

The baseline probability is dramatically different without uncertainty, but the impulse responses are clear: forward-looking plants change their investment behavior to an equivalent change in μ by almost 3 times as much. Thus, dynamic considerations in a plant's stream of misallocation wedges is extremely important in considering their investment behavior.

It is unlikely, however, that changes in μ would be just a one-time event. Given that misallocation may have many persistent elements, it makes sense to consider persistent changes for μ. Calculating these requires more than just a simple shift in the value function. Rather, this assumes that plants optimize under a different dynamic process.

Recall that μ evolves according to the AR(1) process:

$$\mu_{it} = \tau_0 + \tau_1 \mu_{it-1} + \tau_2 ln(cap_{it}) + \tau_3 cc_{it} \tag{1.25}$$

A persistent shock in μ is modeled as an addition to τ_0, the intercept term. New choice probabilities must then be solved under a new value function, EV'.

Table 1.18: Investment Policy Functions in Response to Permanent μ Shock, (500 MW, 2006)

Type of Investment	Baseline	-1 SD	-2 SD	+1 SD	+2 SD
Invest 1000 MW Bin	2.2%	1.7%	1.4%	3%	4.5%
Invest 400 MW Bin	6.1%	5.3%	4.7%	7.2%	9.0%
Invest 0 MW Bin	91.8%	93.0%	93.8%	89.8%	86.5%

Notes: Mean μ is -.05, mean mc is .18 (000 RMB/MWh). Results are averaged over provinces. One SD in μ is .4. Results are generated by solving for new value function with shifted intercept in AR(1) process. Functions are for a Type 2 plant.

Here investment probabilities are drastically altered: a persistent 1 SD change in μ lower a plant's large investment probability by over 20%, and their medium investment probability by over 15%. At 2 standard deviations, both values are near 1/3. In further iterations of this paper, I plan to incorporate truly persistent heterogeneity in μ in my dynamic estimates, rather than just modeling it as an AR1 process. But, these results show that persistent changes in μ are extremely important to plants when making their dynamic investment decisions.

1.6 Counterfactuals

1.6.1 Static Reallocation

The first counterfactual is simple: keeping capacities fixed, eliminate all dispersion in μ. My current results only look at μ, taking the sum of the fixed and random component together. Measuring the gains from reallocation is slightly difficult since production can substitute away or toward the fringe. Thus, an appropriate exercise is to solve for the μ that can be allocated to each plant to keep the fringe's share constant. This μ is equal to

$\mu^a = \frac{(\sum_i \beta_0 + \beta_1 mc_i + \mu_i)}{(\sum_i \beta_0 + \beta_1 mc_i)}$

Overall, the cost savings per unit across the plants included in the sample is about 2.8%. This is likely largely driven by the fact that the largest cost differences are cross-province rather than intra-province. This suggests that there are no easy paths for China to become more efficient, then: while there are gains from reallocation to be had, they will remain overall modest without transmission infrastructure improvements.

We can now reassess the comparison to the US data:

Table 1.19: Correlation Between Utilization and Heat Rate: US vs China by Year

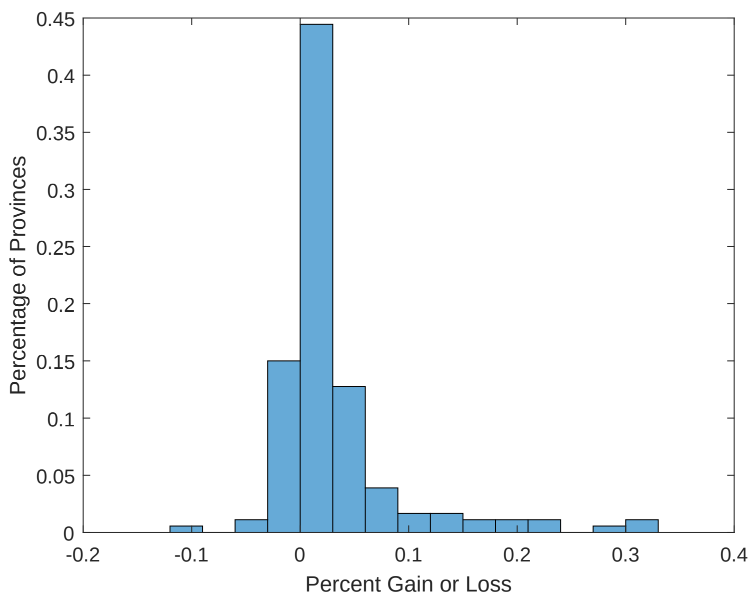
Year	Baseline China	Baseline US	Counterfactual China
2000	-.03	-.34	-.13
2002	.02	-.39	-.14
2003	-.13	-.15	-.21
2004	-.05	-.21	-.23
2005	-.02	-.20	-.21
2006	.09	-.46	-.02

Notes: Results are averaged over provinces. Counterfactual results are generated by eliminating dispersion in policy distortions, but keeping fringe share fixed.

In some years, the reallocation brings China's correlations almost exactly in line with the US, and results in a substantial improvement in almost all years. To some extent, this validates the wedge measurement exercise: efficient reallocation of production does not produce a system that looks totally unlike the current China, or totally unlike a restructured market like the US.

The gains (or losses, which are generated from efficient plants that had too high μs for the estimated value of σ) from shutting the μ's down are heterogeneous by province:

Figure 1.15: Percent Gain in Costs Per MWh Produced Via Reallocation by Province



Notes: Counterfactual results are generated by eliminating dispersion in policy distortions, but keeping fringe share fixed. Regions with net losses are a result of unobserved cost shocks.

As can be seen from this graph, some provinces in some years stand to gain by as much as 20% from reallocation, while others would suffer mild losses. Thus, while the average national level gain is small at around 3%, some provinces would benefit immensely.

Savings are generated by a positive correlation between μ and costs (.33), and a negative correlation between μ and capacities (-.25). That is, planning policies on average are favoring high-cost and smaller plants at the expense of the more efficient large plants. While these gains are modest in percentage terms, in this incomplete sample they still account for over $3 Billion.

1.6.2 Investment

Baseline Counterfactuals

A small change in μ could drastically alter a plant's return to capacity. Counterfactual policy functions suggest that plants will invest very differently in response to their wedge changing.

To gauge this I take a 400 MW plant and start it in the year 2000. Taking 100 draws of types and investment probabilities, I simulate its investment path at the mean value of μ and under a persistent 1 SD decrease.

Table 1.20: Average Simulated Capacity (MW)

Year	Baseline	-1 SD	% Difference
2002	498	479	3.8%
2003	561	516	8.7%
2004	602	553	8.9%
2005	637	583	9.3%
2006	668	609	9.7%
2007	694	634	9.5%

Notes: Results are averaged over provinces. Results are generated by simulating investment paths starting in 2000 for a 400 MW plant at mean cost. 1 SD in μ is .4.

Just seven years out, there is an average 10% reduction in a plant's size. This already generates an over 1.5% difference costs to the plant (due to returns to scale) from the same starting conditions. Combine this with a similar entrant investment exercise:

Table 1.21: Average Simulated Initial Capacity (MW)

Year	Baseline	-1 SD	% Difference
2002	452	373	17.5%

Notes: Results are averaged over provinces. Results are generated by simulating entry decisions in 2002. 1 SD in μ is .4.

The effect for initial investment is even stronger. It is clear that distorting a plant's expected production scheme by 1 SD has a substantial effect on both its expected size and cost. The question remains: can we use the static and dynamic analysis thus far to uncover what may be in the regulators' objective functions in this market?

Concentration Counterfactual

μ being positively correlated with smaller sizes, combined with investment being especially responsive to allocation policies, means that current allocation policy will flatten the plant size distribution. This may be by design: as mentioned before, the exercise of market power is of great concern in restructured electricity markets, as examined in Borenstein et al. (1999) and Borenstein et al. (2002). There is also a long history of government policy in procurement settings to encourage behavior just like this, as documented in Krasnokutskata and Seim (2011), Saini (2011), and Marion (2011), among others.

To characterize the effect of current allocation policies on concentration, I take two representative plants: a 300 MW plant ($\approx$25th percentile) at the observed average μ for that size, and a 900 MW ($\approx$75th percentile) plant at the observed average μ for that size, and simulate their investment paths through 2007.

Table 1.22: Simulated Capacity Paths (MW) at Observed Average μ by Size

Year	300 MW	900 MW	Size Ratio
2002	486	935	1.92
2003	554	947	1.71
2004	579	959	1.66
2005	589	974	1.65
2006	598	985	1.65
2007	619	997	1.61

Notes: Results are averaged over provinces. Results are generated by simulating investment paths starting in 2000 for a 300 MW plant at observed mean μ for 300 MW plants, and a 900 MW plant at observed mean μ for 900 MW plants.

The size ratio of these plant sizes starts at 3, but quickly shrinks as the smaller plant responds to its persistently .high value of μ. I then simulate investment paths swapping the expected μ's for each plant. This would mimic a planner that favors the more efficient, larger plant like the current policy regime is in effect favoring reducing concentration.

Table 1.23: Simulated Capacity Paths (MW) at "Switched" Average μ by Size

Year	300 MW	900 MW	Size Ratio
2002	335	928	2.77
2003	365	991	2.72
2004	391	1022	2.61
2005	415	1052	2.55
2006	436	1075	2.47
2007	453	1111	2.45

Notes: Results are averaged over provinces. Results are generated by simulating investment paths starting in 2000 for a 300 MW plant at observed mean μ for 900 MW plants, and a 900 MW plant at observed mean μ for 300 MW plants, which is reversed from the baseline simulation.

In 2007, the size ratio of the two plants is over 50% larger under the efficiency promoting regime. The disparity between these two values of μ is equal to roughly one standard deviation, so these are not drastically different values. These simulation results, compared with the observed correlations between μ and size, show that the current policy regime is doing a substantial amount of plant size equalization.

1.7 Conclusion

This paper has measured and demonstrated three key aspects of the Chinese coal power industry: In a static context, there is widespread, but potentially modest misallocation of output across plants. Wedges are negatively correlated with capacity, and positively correlated with costs, meaning current planning mechanisms favor smaller, high-cost plants at the expense of larger, more efficient plants. This misallocation comes with significant environmental damage, as there is at least \$90 million in savings from lowered carbon emissions via reallocation, and at higher social costs of carbon this number approaches \$1 billion. These figures refer to a subsample of the full universe of plants that have only 40% installed capacity, and it is likely that they would more than double when extrapolated to every observation.

This misallocation generates significantly different investment behavior among plants than they would engage in without it. Namely, plants that are assumed to be myopic over their policy distortions change their investment behavior by one third as much as forward-looking plants. This, paired with the strong fit of my dynamic model, provides ample evidence both that policy distortions are persistent and plants are aware of this fact. These distortions have significant consequences for the behavior of forward looking plants: a transitory negative 1 standard deviation shock to a representative plant's wedge reduces their investment probabilities by over 5%. When this shock is made permanent, forward looking plants reduce their probability of making a large investment by roughly 20% of the baseline probability. The behavioral changes also generate economically significant changes to lifetime profits: 1 SD transitory shocks change a plant's lifetime earnings by 3-5%, and permanent shocks of the same magnitude change earnings by 60-100%. From a plant's perspective, the amount of misallocation in this industry, while potentially modest in a static model, amplifies greatly in the dynamic context. To ignore the dynamic consequences of these policies is to ignore the vast majority of the influence they have on power plants.

Because power plants tend to exhibit cost savings as they scale up, this altered investment behavior is consequential for costs. In counterfactual simulations, I show a representative plant faces a higher per-unit cost by over 1.5% after just 4 years under a 1 standard deviation lowering of their wedge. This may have substantial environmental implications: if misallocation is inducing suboptimal investment behavior that means more coal is used than necessary, then this already carbon-intensive industry may be emitting more than it needs to. To answer these questions in aggregate requires more of a general equilibrium approach, but this paper provides a crucial empirical first step to answering these questions.

The investment patterns generated by this policy regime are consistent with concerns about market concentration: observed policy distortions are clearly keeping the plant size distribution much flatter than they would be in a planning regime that chiefly promoted cost efficiency. In future work I plan to do market-level simulations establishing some kind of optimal set of wedges for a planner with a clearly defined objective function.

Establishing these findings took several methodological contributions: I developed the first (to my knowledge) capacity investment model that incorporated nonstationarity, unobserved heterogeneity in costs, and perfect foresight over macro shocks together. The incredibly complex regulatory environment in China requires all of these innovations to stay faithful to the real-world setting these power plants are operating in. My approach allowed for a computationally tractable estimation that was able to capture key macroeconomic shocks, market-level heterogeneity, plant-level heterogeneity, and forward-looking behavior from plants over their expected stream of policy distortions.

This analysis also exploited a novel dataset on this industry, and developed a new framework for estimating misallocation in the presence of hard capacity constraints and inter-market electricity transfers. This estimation technique exploits the centrally planned nature of this market, and avoids imposing profit-maximizing production behavior on the power plants. It imposes modest behavioral asummptions on planners themselves. Together, these models allowed me to provide the first empirical investment cost estimates for the largest energy industry in the world. With the additional environmental externalities associated with burning coal, the framework I developed also helps to answer questions that are central to climate change and pollution policy.

This paper sets up a great deal of future work. This work serves as a first step in recovering the preferences of these planners, in a manner similar to Timmins (2002). There are many added complexities in this environment, however: planners have many plants to choose from in allocating any given unit of power production. Thus, in addition to considerations over keeping retail electricity prices low (a likely component of current planning objective functions), I plan to model planners as placing different Pareto weights on each plant, which will help to rationalize both their pricing levels and production.

There is also a vast amount of analysis to be done regarding the notion of private entry under regulatory uncertainty, or where incumbents are politically favored in this context. There are very few papers relating political connections, corruption, and emissions, and incorporating managerial data into model this explicitly would be extremely fruitful. My current dynamic model should also be able to absorb an entry/exit model.

Finally, the wedges recovered from the static analysis should serve as an excellent outcome variable for reduced form and policy analysis in this industry or other developing electricity industries. In a separate paper, I concentrate more specifically on this kind of analysis, and evaluate the 2002 reforms in this market using some of the techniques I have developed in this paper.

CHAPTER 2

INFRASTRUCTURE, TECHNICAL EFFICIENCY, AND MARKET-BASED PRODUCTION: PATHS TO REFORM IN CHINA'S ELECTRICITY GENERATION SECTOR

2.1 Introduction

As of 2010, 78.7% of China's power came from coal, which is almost 40% higher than the global average Liu (2013). Despite current attempts to convert their energy supply over to renewables, China will be dependent on coal power plants in the short term.

Aggregate efficiency improvements in this market have far-reaching welfare consequences. China's coal-powered electricity market is noted to be lacking in several areas: heat rates in Chinese coal fired plants are much higher than those in the US [1], transmission infrastructure is insufficient to transport electricity from coal-rich Western provinces to Eastern population centers, and pricing and production mechanisms are largely planned rather than abiding by conventional competitive principles [2].

China has undertaken serious reforms to address some of these issues, most notably with a large restructuring effort in 2002. In addition to quantifying the outstanding sources of inefficiency in this market, this paper also seeks to assess the impact of these reforms with newly available data on power plants during the period they were undertaken.

My approach starts with two competing cost models: one where the behavior of market authorities is explicitly modeled from Eisenberg (2019), and a linear pure accounting

[1]As established by my data.

[2]Both addressed in Eisenberg (2019).

cost model (with curvature). The former allows me to see if government policy regarding plants changed in response to the 2002 reforms, while the latter allows me to compare efficiency across different possible electricity market sizes to gauge the possible gains from transmission infrastructure improvements. The accounting cost model also serves as a second method to measure misallocation in the market.

To estimate potential intensive margin efficiency gains, I adopt the more reduced form framework of Gao and Van Biesebroeck (2014a), which is itself closely related to Fabrizio et al. (2007). This analysis relies on specifying a cost function for each plant, imposing a cost minimization assumption, and taking a first-order Taylor approximation to specify a regression equation. The policy of interest is then analyzed via a difference-in-differences framework. Gao and Van Biesebroeck (2014a) posit that state-owned firms are more exposed to these reforms and are thus a treatment group. Their analysis relies solely on financial data (and some aggregated control variables), while I can re-run their analysis using pure physical data, which would have been their "preferred" specification. As a preview of my results, I find that the purely physical version of their analysis does not.

This type of analysis is doubly important because China is still rapidly industrializing: China's per-capita energy consumption is still at half that of Western Europe, and a quarter of that of the United States Liu (2013). So, even with substantial conservation efforts, there may be a need to expand China's coal generating capacity, absent major intervention elsewhere in the economy. It is important to do so as efficiently and environmentally sound as possible, to the extent that it is going to happen.

When talking about climate change, there is always a temptation to look to a future where countries like China have all but abolished their coal power plants and replaced them with clean sources like solar, wind, and hydroelectric power. This, of course, is an admirable goal, and one that China is taking seriously in the face of climate change. But, as the numbers above suggest, any serious look at energy and emissions in China has to address costs and benefits on the intensive margin for the coal power industry.

An advantage of studying misallocation in a coal power market is that it is relatively easy to compute plant-level emissions indices. While one can approximate emissions from any kind of manufacturing output, coal use has the advantage of mapping very directly to carbon emissions. It thus becomes very easy to map cross-plant efficiency differences into cross-plant emissions differences, which will be a key point of measuring emissions in any counterfactual re-allocation of resources. Choosing this industry in particular allows for a link between development-style misallocation analyses and carbon emissions that is usually made more indirectly.

A related strain of papers comes from the productivity literature. In papers like Foster et al. (2008), much is made of the distinction between revenue-based and physical measures of productivity. While I am not directly measuring TFP, the main empirical literature on Chinese productivity thus far usually is unable to make this distinction due to data limitations (see, ie Brandt et al. (2014)). My novel dataset will be able to precisely identify prices and quantities for both inputs and outputs, which should lead to a richer analysis of this industry than many prior studies have been capable of.

Asker et al. (2017) undertake a similar analysis in the global oil market to determine OPEC's role as a cartel in contributing to misallocation. They, too, structurally estimate each firm's cost function and calculate a socially optimal aggregate solution using a "sorting algorithm." This is in effect the same as ordering dispatch based on marginal cost. They then run several counterfactuals holding various aggregate production quantities fixed.

As mentioned before, a prominent, recent study on the restructuring of electricity markets is Fabrizio et al. (2007). Studying the US, the authors find "modest medium-term efficiency benefits from replacing regulated monopoly with a market-based industry structure." In England and Wales, Newbery and Pollitt (1997) find modest increases in in efficiency as well, but very minor benefits for consumers. China's radically different market makes direct comparisons difficult, but these studies have established baseline empirical methods for studying this industry.

Some of the more negative effects of electricity restructuring–which may help to explain why China has continued its current regulatory regime–have been seen in California's electricity market. Papers like Borenstein et al. (2002) and Borenstein et al. (2008) have helped shed light on the various incentives in such a market and their consequences. These considerations are worth keeping in mind as China possibly moves toward this kind of market, but the best data available is only at the yearly level and is very coarse. At any rate, even under reforms that would bring about more market competition, China likely intends to keep substantial central control over the industry Liu (2013).

Finally, the potential gains from infrastructure improvements in developing countries have been studied recently by Ryan (2014) in India, though due to data limitations my paper is forced to take a more abstract approach.

2.2 Industry Background

In April 2017, a report from Resources for the Future claimed that "China currently does not have a spot market for electricity" Ho et al. (2017). This is because production is largely allocated, and prices are tightly controlled. The guiding allocative principle is for each plant to have roughly the same number of operating hours, though this is selectively enforced and highly variable across provinces. As the RFF report makes clear: Given that generation planning is a decision at the provincial level, it should be expected that different provinces will dispatch generators differently, and this heterogeneity should be properly accounted for." Additionally, the report claims that the rationale behind this type of allocation is "primarily distributional."

Incentives for plants to be efficient are limited in this context. If they cannot really influence prices and will be given similar amounts of production regardless of what they do, it stands to reason that they will not make costly investments in streamlining their production processes. Similarly, plants with existing efficiency advantages are unable to exploit them due to this allocated production model.

China's current production model began around 1998. 1998 marked the start of a shift toward efficiency-focused reforms after decades of the government largely trying to increase capacity to meet power demands according to Xu and Chen (2006). The state no longer held a total monopoly over the power generation industry like it traditionally had, and now had "a market structure composed of diversified investors" Xu and Chen (2006). However, this by no means resulted in a smoothly-functioning market system. As Xu and Chen (2006) state: "The reform in the electricity industry was mainly on the governmental level, the old regulatory system did not change at all in the lower levels, which remained incompatible with both the power industry's market-oriented reform and diversified operating entities...influence from the central government was still very large and the governments, both central and regional, played an important role in the industry. A modern regulatory system was far from coming into being."

Put differently, the central, regional, and provincial governments all still played (sometimes conflicting) roles in a plant's operation. These actors often had differing political and economic objectives: a provincial head would likely care about maximizing province-level output or profits rather than ensuring a more efficient allocation of resources across a wider geographical areas. This is especially important in China, where coal resources are not evenly distributed across the country. As Xu and Chen (2006) put it: "Areas rich in primary energy deposits were far from power-load centers. However, market segmentation by administrative divisions exerted a tremendous impact upon resource allocation; power from cheap, clean energy sources were rarely distributed across provincial divides due to inter-political barriers." Only adding to these frictions is China's underdeveloped transmission apparatus, which adds a physical barrier to the existing political ones.

Xu and Chen (2006) also indict the pricing system, claiming there was both a lack of uniformity and enforcement across plants. As a result, "prices could not reflect the true relationship between supply and demand." The authors also claim that pricing and investment conditions places independent power producers "at a disadvantage compared to state power plants" at times when there was enough capacity to meet demand.

In 2002, several major reforms were enacted. They involved breaking up a major state-owned enterprise into five smaller companies and separating administrative functions at the federal level for transmission and generation. There was also a contemporaneous deregulation of the input (coal) market [3].

The literature suggests that many of these problems persisted well after 2002: Liu et al. (2013), writing in 2013, say that "power-generating companies...must sell their output at regulated prices that often do not cover costs." Massive financial issues arose around 2010, and in 2011 "the top five state-owned power generation groups lost more than $1.5 billion on their thermal power operations in the first quarter of 2011." Qual-

[3] See, ie, Gao and Van Biesebroeck (2014a) for more information)

itatively, it seems that these initial reforms have done little to resolve this industry's problems. Gao and Van Biesebroeck (2014b) use the methodology from Fabrizio et al. (2007) to assess whether the 2002 "restructuring" of this market lead to intensive-margin efficiency gains. While they find modest gains, they also face data availability issues that may prevent their results from being fully conclusive.

2.3 Data

The key dataset to this paper is a confidential survey of coal power plants conducted by the Chinese government. It covers, roughly, the universe of power plants from 1997-1998, 2000, and 2002-2011 [4]. Major variables include a plant's name, power generated, coal used, and nameplate capacity. The plant's name allows us to find locations and ownership status–the latter is extremely important for determining which plants were and are owned by the "big 5" state-run corporations, as well as plants that are owned partially by the state. The fullest version of the dataset contains 21,121 plant-year observations. From this data I can also derive a plant's "heat rate", a standard measure of efficiency calculated by dividing coal input by power output. This will be the main index I use to assess cross-plant physical efficiency levels, and the associated emissions from each plant's output in counterfactual scenarios.

A subset of these observations are then merged with the now-standard NBS census data from 1998-2007 to get financial information. This includes a plant's revenue, various accounting cost measures, material expenditures, location, state ownership status, capital stock, investment, and employment information. These variables are inconsis-

[4]Thank you to Shanjun Li, Deyu Rao, and many others for preparing this data and allowing me to access it.

tently kept across different years, so the sample varies depending on one's analysis (for an example, plants with non-missing revenue information are about 4,200). For these observations I can obtain output and input price indices, which will prove useful in distinguishing between financial measures of misallocation and physical ones.

My estimation requires using weather data as an instrument, which I get from the NOAA's land-based station data. Using geographical coordinates, weather stations are matched to the nearest county. For the observations that I am able to merge with the financial census, I can identify which county they are in, and thus I can get weather data at a sub-province level. Key weather variables I use are average temperature, average minimum temperature, dew point (for humidity), and visibility.

Finally, I use the CEIC database for certain province-level indices–they provide retail electricity prices, gas prices, average coal prices, and some aggregate transmission data. While I have not fully incorporated this data into my analysis yet, it has been serving as a useful check for some of the figures I have been generating.

2.3.1 Summary Statistics

All variables (except N) are means. The summary statistics show that the market is growing both in terms of number of firms, average firm size, and production. Firms are also getting slowly more efficient over time. Note that input prices are only available through 2007.

Table 2.1: Means of Major Variables, 1998-2007

Year	Cap (MW)	Prod (MW)	Price (000 RMB/MWh)	Phys. Cost (000 RMB/MWh)	Heat Rate (tons/MWh)	N
1998	384.58	216.73	0.26	0.17	0.55	195
2000	422.75	239.15	0.28	0.17	0.60	193
2002	484.50	297.51	0.27	0.17	0.60	221
2003	491.99	328.71	0.29	0.19	0.55	232
2004	509.97	349.23	0.26	0.19	0.56	283
2005	549.39	361.71	0.29	0.23	0.56	292
2006	636.19	391.38	0.28	0.22	0.54	326
2007	693.14	416.70	0.31	0.25	0.53	351

Notes: Table depicts summary statistics for years 1998-2007. Physical variables are from confidential power plant survey, financial variables are from a combination of physical dataset and financial variables from annual NBS manufacturing census. One RMB is roughly .15 dollars, so the output price in 1998 of .26 000 RMB/MWh would equal about 40 dollars per MWh, while the 2007 output price would be more like 47 dollars. Figures are for sample where revenue and physical data is matched.

2.4 Suggestive Results

2.4.1 Costs and Misallocation

It may be that these reforms were executed poorly, but it may also be the case that it is difficult for plants to improve their physical efficiency in a short timeframe in China. As such, it is worth considering whether there are better improvements to be made by prioritizing different plants and allocating them more inputs based on efficiency. Asker et al. (2017) undertake a similar exercise, and with an eye toward quantifying aggregate misallocation measures, it is necessary to estimate a slightly different structural cost function for each plant.

With that baseline established, we can delve more deeply into the sources of China's misallocation.

Note that columns 2 through 4 include province fixed-effects. We can see from the above table that heat rate is only significantly related to utilization in the simplest

Table 2.2

Descriptive Regressions - Heat Rate, 2000

VARIABLES	(1) Utilization	(2) Utilization	(3) Utilization	(4) Utilization	(5) Utilization
Heat Rate	-0.0401	-0.00233	-0.00122	0.0541	-0.0173
	(0.0402)	(0.0401)	(0.0404)	(0.0610)	(0.0710)
Capacity	-3.23e-05	-6.16e-05***	-6.23e-05***	-9.93e-05***	-0.000102***
	(2.21e-05)	(2.07e-05)	(2.09e-05)	(2.62e-05)	(2.59e-05)
Entry	-0.0653***	-0.0726***	-0.0722***	-0.0988***	-0.0947***
	(0.0182)	(0.0181)	(0.0183)	(0.0257)	(0.0255)
Big 5			0.00502	0.0110	0.0107
			(0.0186)	(0.0212)	(0.0209)
Output Price				-0.0380	0.121
				(0.0610)	(0.103)
Input Price					-0.174*
					(0.0914)
Constant	0.616***	0.722***	0.719***	0.835***	0.889***
	(0.0283)	(0.0639)	(0.0653)	(0.118)	(0.120)
Observations	340	340	340	149	149
R-squared	0.044	0.311	0.312	0.580	0.593

specification–once we control for province, the effect completely disappears. Given that provinces have some level of governing authority over their electric grids, and different provinces may target different mean utilizations, it is probably appropriate to weigh specifications with the fixed effects more heavily.

The sign on the entry coefficient is unsurprising–plants that have just entered are likely to operate for less of the year than incumbents. Of interest are the two price variables (output and coal). We can see that lower input prices are associated with significantly lower utilizations–it is hard to interpret this result precisely, since there are two countervailing forces at work here: Lower input prices could reflect worse quality coal inputs, which could make operating profitably more costly and lead to lower

Table 2.3

Descriptive Regressions - MC, 2000

VARIABLES	(1) Utilization	(2) Utilization	(3) Utilization	(4) Utilization
MC	-0.169	-0.122	-0.117	-0.373*
	(0.102)	(0.0920)	(0.0933)	(0.218)
Capacity	-6.90e-05**	-0.000112***	-0.000113***	-0.000116***
	(2.90e-05)	(2.34e-05)	(2.37e-05)	(2.37e-05)
Entry	-0.0773***	-0.0997***	-0.0994***	-0.101***
	(0.0288)	(0.0254)	(0.0255)	(0.0254)
Big 5			0.00744	0.00801
			(0.0210)	(0.0209)
Output Price				0.183
				(0.141)
Constant	0.655***	0.877***	0.877***	0.874***
	(0.0286)	(0.112)	(0.112)	(0.112)
Observations	149	149	149	149
R-squared	0.085	0.581	0.581	0.587

utilizations. On the other hand, lower input costs should mean a plant is operating at a lower marginal cost, which would incentivize them to have a higher utilization than their competitors. At any rate, not a single one of these specifications indicates that lower heat rates lead to higher utilizations, controlling for many salient factors.

Again, specifications starting with column 2 include province fixed effects. The story changes somewhat when we look at pure marginal cost, which is the product of a plant's heat rate and input price. Lower marginal costs are associated with higher utilizations, so in this sense more efficient plants seem to be producing more.

In 2006, when plants faced a **more** competitive input market, this effect has all but disappeared. This suggests that output prices and/or designated production quantities are not allowing firms to leverage cost advantages they have against each other. In

Table 2.4

Descriptive Regressions - MC, 2006

VARIABLES	(1) Utilization	(2) Utilization	(3) Utilization	(4) Utilization
MC	-0.133**	-0.168***	-0.172***	-0.434***
	(0.0605)	(0.0622)	(0.0627)	(0.161)
Capacity	-5.76e-05***	-7.63e-05***	-7.42e-05***	-7.77e-05***
	(1.35e-05)	(1.38e-05)	(1.43e-05)	(1.44e-05)
Entry	-0.131***	-0.133***	-0.134***	-0.131***
	(0.0224)	(0.0232)	(0.0232)	(0.0232)
Big 5			-0.00959	-0.00666
			(0.0164)	(0.0164)
Output Price				0.238*
				(0.135)
Constant	0.738***	0.860***	0.867***	0.847***
	(0.0205)	(0.0677)	(0.0687)	(0.0694)
Observations	334	334	334	334
R-squared	0.148	0.298	0.299	0.306

fact, the regulatory regime in place before the coal market was deregulated around 2002 seems to have been ensuring a more efficient allocation of inputs based on this measure.

Capacity is highly significant in all of these regressions, and the results are highly sensitive to its inclusion, suggesting it is important to look at the size distribution of plants in China, and the relative utilizations across this distribution.

2.4.2 Infrastructure

Some plausible metrics to examine potential gains from transmission infrastructure before applying any kind of model would be regional input prices and heat rates. This should give us a sense of how much cheaper or more efficient the western regions actually are.

Table 2.5: Input Prices (000 RMB/MWh) for Representative Regional Grids

Year	Northwest	North	Northeast
1998	.211	.54	0.33
2003	.26	.34	0.38
2007	.64	.45	0.44

Notes: NW includes Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang, Tibet. N includes Beijing, Tianjin, Hebei, Shanxi, Shandong. NE includes Liaoning, Jilin, Heilongjiang.

Table 2.6: Heat Rates (tons/MWh) for Representative Regional Grids

Year	Northwest	North	Northeast
1998	.58	.58	0.58
2003	.60	.60	0.55
2007	.51	.62	0.55

Notes: NW includes Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang, Tibet. N includes Beijing, Tianjin, Hebei, Shanxi, Shandong. NE includes Liaoning, Jilin, Heilongjiang.

We can see for earlier years in the sample that the western, coal rich regions are substantially cheaper, with roughly similar heat rates. By 2007 it appears that there has been a policy change that increased coal prices more in the west than it did in the east. These prices and heat rates are endogenous to how coal markets operate, so these tables do not tell the full story, but it appears that building high powered infrastructure may lead to more muted gains than the coal stores of China would initially suggest. Later in the paper I will examine this in full using plant-by-plant measures and accounting for capacity constraints.

2.5 Model

There are three related models in this paper that will be used to examine the market from different angles: a cost model borrowed from Gao and Van Biesebroeck (2014a) and Fabrizio et al. (2007) to examine cost responses to restructuring using a difference-in-differences framework, an model of accounting costs to look at possible gains from reallocation without a pure structural model imposed, and a model of how policy planners in China allocate production based on preferences and unobserved costs, borrowed from Eisenberg (2019).

2.5.1 Planner Behavior Model

The bulk of this model is derived in Eisenberg (2019), so I present an abridged version here. This model differs from the other two in that it explicitly accounts for a planner's preferences and estimates their policy function for assigning production to different power plants. While this results in a stronger set of assumptions, it allows me to measure how production is being allocated in China, since this particular market does not abide by any traditional competitive framework.

The gist of the model is that a planner who makes a continuum of small decisions about how to allocate production across each power plant can be represented by the following equation:

$$ln(\frac{q_{it}}{cap_{it}}) - ln(\frac{q_{0t}}{c_{0t}}) = \beta_0 + \mu_{it} - \beta_1 cc_{it} \tag{2.1}$$

Where q represents plant's allocated production, index 0 represents a normalized "fringe" firm for each (province-level) market, and cc represents a plant's linear coal cost, observed directly in the data. μ represents the net effect of a planner's other goals/preferences, while β_1 controls for possible unobserved cost shocks or curvature in a plant's cost function.

2.5.2 Gao and Van Biesebroeck (2014a) Cost Model

Gao and Van Biesebroeck (2014a) borrow the cost-minimizing estimation framework from Fabrizio et al. (2007) to test the results of restructuring. The full derivation of their technique is available in those papers, but I will provide a basic outline of the estimation to explain my replication results.

The key identifying assumptions are that firms have a CES production function and are cost-minimizing. After some algebra, a first-order Taylor approximation, and consolidation of terms, these result in the following basic log-log models:

$$lnM_{it} = \gamma_i + \gamma_t + \gamma_1 lnQ_{it} + \varepsilon_{it}^{M} \tag{2.2}$$

Where M is a firm's material inputs, i indexes firms, t indexes years, and Q is a firm's output. This provides a basic regression framework to identify the effects of restructuring. The authors argue that state-owned firms were more exposed to restructuring than private ones, and thus can be thought of as a "treated" group for use in a differences-in-differences framework. While it would be ideal to combine the DiD framework with the simple physical demand equation above, Gao and Van Biesebroeck (2014a) only have access to financial data, and also have to account for missing prices:

$$lnM_{it} = \gamma_i + \gamma_t + \gamma_1 lnQ_{it} + \gamma_P X_{it} + \mu_t STATE_{0i} * Restruc_{it} + \varepsilon_{it}^{M} \tag{2.3}$$

$$lnEMP_{it} = \rho_i + \rho_t + \rho_1 lnQ_{it} + \rho_W lnWAGE_{it} + \rho_P X_{it} + \zeta_t STATE_{0i} * Restruc_{it} + \varepsilon_{it}^{L} \tag{2.4}$$

X_{it} is a set of missing price controls, including firm size, and measures like firm size interacted with province dummies. μ_t is the parameter of interest, which represents the interaction between state ownership and post-restructuring. The findings are robust to using either time fixed effects or a simple indicator variable for pre- and post- wherever time shows up in this equation, both in the original paper and my specification.

A negative sign on μ would indicate that restructuring caused firms to become more efficient. Material use would have declined in response to the policy for treated firms, holding the level of output fixed. This would represent intensive margin efficiency gains, where plants themselves became physically more efficient in response to new incentives.

The centerpiece of this cross-plant analysis is a firm's cost function. I implicitly assume that firms are cost-minimizing, and estimate relevant parameters from the financial and physical data that I have. When estimating cost functions of this form, papers like Fowlie et al. (2016) assume that firms are profit maximizing, and use optimality conditions to identify cost parameters. I cannot make this assertion, because it is well documented that firms are not in control of their prices, and that profit-maximization is at best only one of many conflicting factors determining a firm's output decision.

2.5.3 Accounting Cost Model

A thermal plant's costs are roughly 80% fuel-based, and are generally very close to linear (see, ie Ryan (2012))). If I am able to capture the difference in variable fuel costs across plants, this should be the first-order difference in overall costs as well. Additionally, since these plants are chiefly using just one material input to produce one output, I can plausibly fit an accounting variable cost measure like cost of goods sold to production data and generate results that are capturing the economic concepts of marginal and total cost. An advantage to this approach is that it requires relatively few structural assumptions on how policymakers operate in China or plants behave, at the cost of having little power to capture costs that do not show up directly in accounting data.

Working only from cost data, if there is curvature to a plant's cost function I can generate an optimality condition given estimated cost parameters, and compare what firms actually produced to what would have been profit-maximizing. This difference will form a "wedge", common to the misallocation literature, though it differs from those I develop later in the paper.

I specify a plant's accounting cost function the following way:

$$cost_{imt} = \phi_0 + \phi_c cap_{imt} + \phi_{1m} cc_{imt} q_{imt} + \phi_2 1(u_{imt} > \eta)(q_{imt} - \eta cap_{imt})^2 + \alpha_t + \phi_{ent} ent_{imt} + \varepsilon^a_{imt} \tag{2.5}$$

Observations are at the plant-year level, where i indexes a plant, m indexes a market, and t indexes a year. $cost_{imt}$ represents a firm's total operating costs, generally measured as COGS or as operating costs (the accounting system changes from year to year, and I have done my best attempt to harmonize them). ϕ_0 and $\phi_c cap_{imt}$ are meant to represent a firm's fixed costs of production. The idea is that a firm incurs some cost just to "turn on", in addition to its fuel costs. cap_{imt} represents a plant's nameplate generating capacity, and the fixed costs scale with this to account for the complexity of larger plants and units.

cc_{imt} is a measure of a plant's physical marginal cost. This is the price of coal multiplied by a plant's heat rate. Keeping everything in terms of megawatt-hours (mWh), I calculate the heat rate by dividing a plant's coal input by its power output, and I calculate a plant-level coal price by dividing its material expenditures by its coal input. Material expenditures are sparsely recorded, and this limits the sample for my full analysis significantly. Environmental regulations and taxes would affect cc_{imt} as well, but there is no easily available information on what sorts of tariffs would hit each firm, so some of this will likely be absorbed by the "wedges" that I analyze.

q_{imt} is a plant's physical output, which I observe directly. ϕ_1 would be exactly equal to 1 in an ideal world–the additional cost of producing a unit of electricity is exactly equal to the cost of buying it and burning it–but I allow it to be estimated and to vary by province. I do this for a few reasons: First, there are unobserved cost factors that, while possibly very small, may be reflected in the accounting cost measures. For the most part, these should be at least proportional to output [5]. Second, it's possible that these costs differ by geographical location or governing authority (ie, maybe there is some unobserved transport cost in denser areas, or as you move further east). Third, letting β_1 be estimated serves as a sanity check. If it deviates to far from 1 in any province, I will know my model is likely badly misspecified.

u_{imt} is a firm's utilization, which is the ratio of their quantity produced to their nameplate capacity in mWh. It thus can take on values between 0 and 1. The ϕ_2 term is meant to represent a plant's capacity constraint, as in Ryan (2012). η represents a certain threshold beyond which a plant's total cost increases quadratically. The idea is that costs are generally linear in output, but as more and more of a plant's resources are used things become costly (like maintenance or overtime).

I include year fixed effects to capture unobserved cost trends (ie, maybe the prices of other inputs are changin). I also include an entry dummy for two reasons: First, it may be more costly to begin operations if a plant is just opening. Second, since I only

[5] An example of this would be wages, which I actually observe directly for these plants. I have tried to omit them in this analysis for simplicity's sake.

observe data at the yearly level, it may be that a plant opened at some point in the middle of the year. This should lead to lower total operating costs, and this dummy helps me to capture them. It may also be the case that entrants have more advanced technology on average than incumbent firms. Thus, there is no obvious sign that the entry parameter should take.

Once these parameters are estimated, which I outline below, we can start to estimate deviations from optimal output. While there may be some avenues for plants to change their prices on their own, for the most part prices are set exogenously (see, ie, Xu and Chen (2006)). Thus, profit maximizing conditions are taken treating each plant as a price taker.

If a plant wants to produce at all (that is, total revenue can exceed $\phi_0 + \phi_c cap_{imt}$), it will produce into the convex portion of its cost curve. This is simple enough to reason out: since the plant is a price taker, $MR_{imt} = p_{imt}$, and via assumption this does not change as they produce different amounts. If $p_{imt} < \phi_{1m}$, then no unit of electricity will ever be profitable to produce. If $p_{imt} \geq \phi_{1m}$, then it will be profitable to keep producing at least into the convex portion of the cost curve, since before then marginal cost is constant. Thus, firms will want to not produce at all, produce optimally at an interior solution at a utilization level past η, or produce at capacity (or as close as possible).

Firms producing in the interior are thus solving:

$$\max_{q_{imt}} p_{imt} q_{imt} - \phi_0 - \phi_c c_{imt} - \phi_{1m} cc_{imt} q_{imt} - \beta_2 (q_{imt} - \eta c_{imt})^2 \tag{2.6}$$

This results in the following FOC:

$$p_{imt} - \phi_{1m}cc_{imt} - 2\beta_2(q_{imt} - \eta * cap_{imt}) * (\frac{1}{R}) = 0 \tag{2.7}$$

Where R is the factor that converts capacity in mw to capacity in mwH. This equation is easily solved in terms of q_{imt} (embedded within u_{imt}) to get a plant's optimal interior production level q^i_{imt}. A plant's profit-maximizing choice of quantity, q^*_{imt} will thus be the discrete choice between 0, q^i_{imt}, and $c_{imt} * R$.

The degree of misallocation for each plant is the difference between optimal and observed outputs, $q^*_{imt} - q_{imt} = \tilde{q}_{imt}$. This can also be normalized by a plant's capacity, which I will denote as τ_{imt}.

2.6 Estimation

2.6.1 Planner Behavior Model

As established in Eisenberg (2019), this model can be estimated using OLS or IV regression as it is linear in parameters.

2.6.2 Gao and Van Biesebroeck (2014a) Cost Model

This model is also linear in parameters. It can be estimated using OLS or IV as well. In the original paper, the authors instrument revenues using market revenue, and I follow suit.

2.6.3 Accounting Costs

There are two complications in my estimating equation that prevent me from doing straightforward OLS. The first is that quantity decisions could easily be correlated with the error term. While TFP concerns outside of a plant's heat rate are hopefully secondary in this market, unobserved components of productivity could still hamper my estimates. Similarly, since these quantity decisions are actually made continuously throughout the year, it's possible there were some persistent shocks that plants were making their decisions with the ability to forecast in the short-term. Thus, I will resort to finding instruments to estimate this equation.

The second concern is that η is nonlinear. It cannot be estimated in a standard way, and its inference is even more complicated. To resolve this, I borrow both estimation and inference techniques from Hansen (2017). The estimation technique is fairly straightforward: pick a grid of points and run the appropriate regression model at each point, and then pick the point that minimizes the criterion function overall. If we write the regression function as a function of η, $\beta' x_t(\eta)$, the estimator can be written the following way:

$$(\hat{\beta}, \hat{\eta}) = arg \min_{\beta, \eta} S_n(\beta, \eta) \tag{2.8}$$

Where S_n is the appropriate criterion function (in this case it will be a GMM estimator, discussed below).

Inference (which I have not done yet) is slightly more complex: "the estimates of the regression function itself are not asymptotically normal...since the regression function is a nondifferentiable function of the parameter estimates. Consequently, conventional inference methods cannot be applied to the regression function," Hansen (2017). To test for linearity and generate confidence intervals, asymptotically valid p-values can be calculated using the bootstrap to approximate the limiting distribution of the F statistic generated by comparing the models with and without the regression kink.

Because I am using instrumental variables, the criterion function will not be straightforward OLS. Determining this function will obviously require specifying which exact instruments I am using. To start, both $MC_{imt}q_{imt}$ and u_{imt} will be endogenous, so at least two instruments are necessary. Additionally, β_{1m} varies by market, so key instruments will also need to be interacted with province indicators.

To start, I am using weather variables from the NOAA as instruments. I can match them at the county level to many of my observations that are in the financial census, and the panel covers all the years necessary. The relationship between weather demand and electricity, while reasonably well understood, is not straightforward. For example. according to Hor et al. (2005), there is clearly a seasonal pattern in electricity demand that relates to temperature, but there is a "nonlinear dependence of demand on temperature at the hot and cold temperature extremes." They also find that including variables related to humidity in their model greatly improves their forecasts.

As a result, I choose four variables from the NOAA data: average temperature, average minimum temperature, visibility, and a humidity index (which I calculate by subtracting the dew point from the average temperature). For now I just include linear terms for all four, and I also interact each of them with province indicators. The model is thus overidentified. Eventually, I plan to do some kind of more thorough model selection procedure (maybe LASSO) to figure out the best possible functional form.

2.7 Parameter Results

2.7.1 Planner Behavior Model

Current results utilize lagged cost and total investment for the five largest firms in a plant's market as instruments:

Table 2.7: Allocation Model Estimates of $1/\sigma$

	(1)	(2)
Estimate	1.61***	.235***
SE	(.594)	(.094)
Individual FEs	Yes	No
Year FEs	Yes	Yes
Province FEs	No	Yes
First Stage F	18.01	567
J Stat p-value	.30	.07
N	1501	1501

Notes: σ is variance of unobserved cost shocks, or the coefficient on marginal cost. $1/\sigma -- > 0$ would imply perfect merit-order dispatch. Dependent variable is log utilization minus log fringe utilization.

An estimate for σ of around .62 implies a substantial role for unobserved cost shocks. Given an average marginal cost of around .19, intra-annual demand considerations are clearly a large factor in determining which plants get allocated production. This will bear out significantly in the results comparing predictions from this model to the structural model of planner behavior.

Table 2.8

OLS Results - Threshold at 0

	(1)
VARIABLES	Cost
Capacity	90.05***
	(33.67)
Entry	-33,990*
	(17,804)
Threshold	0.154***
	(0.0311)
Constant	98,042***
	(26,025)
Observations	1,943
R-squared	0.841

2.7.2 Accounting Cost Model

OLS

Running a grid search for η over an OLS version of the estimating equation gives 0 as the RSS-minimizing quantity, with the following regression results:

There are some encouraging aspects to this result. The R-squared is fairly high, which suggests there are not many other costs being missed. The coefficient on capacity is significant and positive, but not overwhelmingly large. For reference, the mean of the dependent variable is 628,683 (which I believe is in 1,000 RMB, so this is approximately \$94 million) This suggests that there are meaningful fixed costs to operating that do scale with size, but they are generally much smaller than the actual operating costs of the plant. The mean capacity in the sample is about 520 MW, which would make for a fixed cost of about 50,800. So, this model, with all its endogeneity problems, estimates fixed costs to be about 1/12 of the total costs of the plant.

While I have not reported them in this table for brevity's sake, the coefficients on $MC_{imt}q_{imt}$ are close to 1 for just about every province. The smallest for any province is .86 in Inner Mongolia, and the highest is 1.53 in Hebei [6]. The inclusion of wages does not meaningfully shift the R-squared or the significance/sign of any major coefficients.

We can see that the estimate for the effect of being an entrant is significantly negative–this implies the censoring and technological effects win out over the costs of beginning operation. It is possible that the latter costs are largely contained in other accounting variables.

0 is not an encouraging result for η. The result for other, similar industries in the US tends to be around .8 Ryan (2012). There is good reason to suspect that this number would be lower in China due to different labor and technological conditions, but an estimate of 0 does not correspond to the traditional "hockey stick" model. Rather than increasing rapidly above some threshold, this model estimates costs to be gradually decreasing along the entire utilization curve. This isn't necessarily implausible–but the two types of cost increase are not mutually exclusive, and suggest a more complex model may be necessary.

[6] I do not have extremely detailed province-by-province explanations for these results yet, but these make a fair amount of sense. Inner Mongolia is relatively sparsely populated, and has many coal reserves. Wages and transport costs are likely to be lower, while Hebei is more densely populated and is further east, and likely contains a lot of power plants that serve Beijing and Tianjin.

IV

Accounting for data that have weather observations available, the sample drops to 1,847 when the IV is used. With the full set of weather instruments, the optimal η threshold becomes .23:

Table 2.9: IV Results - Threshold Estimated at .82

VARIABLES	(1) Cost
Fuel Costs	2.103**
	(1.059)
Threshold	-1,748
	(1,101)
Capacity	-574.9
	(807.2)
Entry	55,705
	(106,986)
Constant	519,789
	(412,730)
N	1,818
R-squared	0.095

Instruments include humidity, minimum temperature, and average temperature. The threshold of .82 is estimated using grid search over values to find the lowest GMM criterion function.

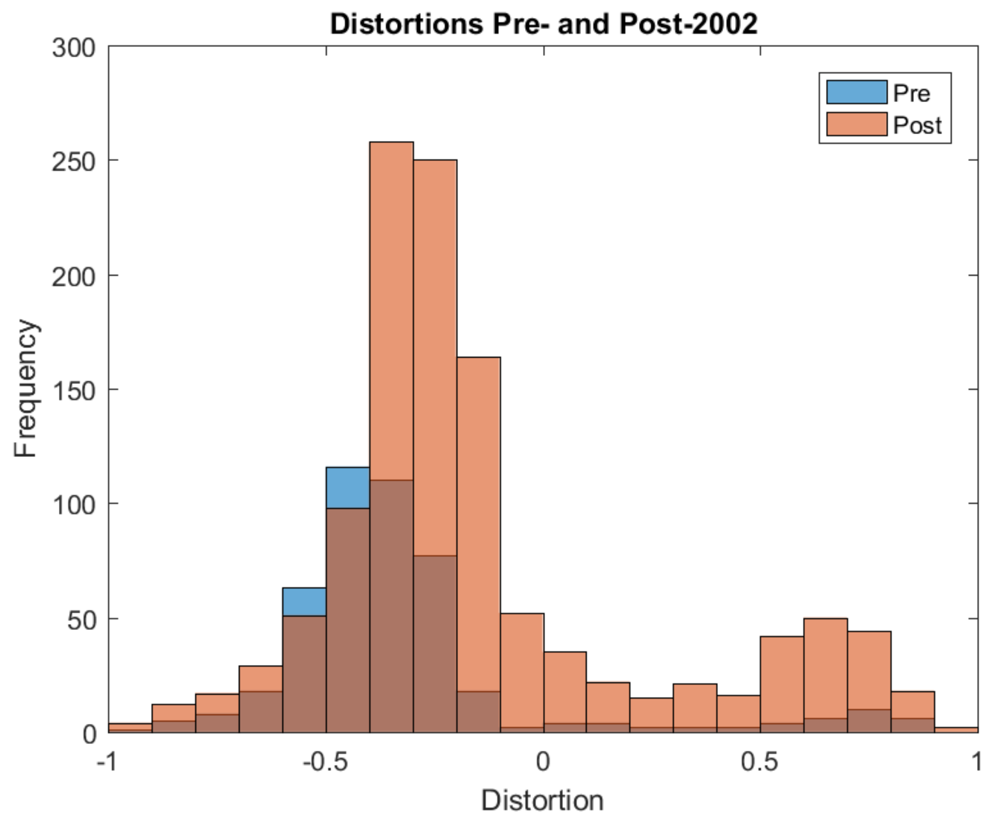
Note that there is only one weather station in the data in Hainan and Tibet (which I should probably exclude anyway) so these provinces have been omitted. The fixed cost estimates are largely unchanged, though somewhat smaller now. The coefficient on the threshold variable has gone up significantly, but because the threshold is now so much higher, an increased utilization is now associated with a substantially lower cost increase than in the OLS specification. A firm at the 75th percentile of utilization only sees an increase of about 1000, which is less than the increase at the overall mean utilization in the last specification.

The entry effect is largely the same in this specification, and the constant cost estimate is substantially smaller. Fixed costs are no longer statistically significant (and the point estimate is in fact negative, though paired with the constant many smaller firms would still face positive fixed costs).

The costs of passing the threshold are substantially higher now, which make sense given that the threshold has also increased. For reference, the median of the dependent variable is 410390 ($61.5 million). This plant's utilization is .55, and it has a capacity of 469 MW. The additional costs incurred by crossing the threshold are 6,081 (again, in 1,000 RMB, so this would convert to about $1 million), which are more or less negligible (around 1.5%) compared to the total costs. These costs start to get more onerous toward the upper end of the utilization curve (which is the point of including them). If this same plant had around 95% utilization with the same linear marginal cost as before, then "threshold costs" would account for almost 6% of total costs. In an industry where everyone is selling a homogeneous product, this kind of cost difference could be a serious factor in determining profitability. Also, when consideration efficient input reallocation, it is likely we will want to take the most efficient plants and get their utilization as close to 1 as possible, so these threshold costs may be significant in welfare calculations.

.23 is likely low for the actual threshold when maintenance costs kick in for a power plant, but we can see from the calculations above that the estimates lead to much more onerous costs at extremely high utilizations. This is likely a combination of two factors: uncertainty around the estimate of η, and slight model misspecification. The former can be addressed somewhat with formal inference, which will come later. The latter may require trying more complex models. It is plausible that utilization has both an unspecified smooth relationship AND a threshold effect with costs, which will be significantly

Figure 2.1



harder to estimate precisely. It may also just be that a fuller set of controls is necessary and the threshold estimate is picking up other costs.

Distortions from Accounting Costs

As outlined in the model above, these estimates imply a set of distortions, which I have normalized by each plant's capacity to get a measure from 0 to 1:

Distortions are calculated by subtracting the optimal quantity from the observed quantity, so a negative number represents a firm that was made to produce less than was optimal from a profit-maximizing perspective. A market where all firms were producing optimally would have all observations at 0, while a "basically" optimal market with some measurement errors and unobserved minor frictions would likely be roughly normally distributed around 0, possibly with a slight skew depending on the most prevalent frictions.

This distribution means that most plants are being made to produce less than is profitable, which suggests that regulators are allocating quantity to less efficient plants out of distributional concerns. It appears that restructuring did not meaningfully alter this pattern, though there now appears to be a large right tail of plants being forced to produce above what is profitable.

2.7.3 Gao and Van Biesebroeck (2014a) Cost Model

Replication

The contemporaneous revenue variable may have endoegneity issues, so I instrument it with province-level production. The results are also robust to using twice-lagged revenue, though the first stage F statistic is less than 12 in the materials equation, suggesting this is not a strong enough instrument. Province-level production results in a first-stage F of about 20. All first-stage F's are well above 12 for the labor equations.

Table 2.10: Gao and von Biesebroeck Specification Using Revenues

VARIABLES	(1) Log Coal	(2) Log Employment
Log Revenue	0.074	-0.2
	(0.084)	(0.13)
Log Employment	0.086	
	(0.20)	
Age	0.0002	0.0001
	(0.0002)	(0.0001)
Restruc	0.44***	-0.01
	(0.0650)	(0.0599)
Restruc x SOE	-0.207***	-0.111**
	(0.0654)	(0.0542)
Log Wage		-0.529
		(0.380)
Constant	10.01***	6.919***
	(0.935)	(1.128)
N	2,634	1,822
Plants	731	654

Notes: Revenues are instrumented using market revenues. Regressions include province fixed effects interacted with log employment and individual fixed effects.

Both the material and labor equations show that the effects of restructuring are significant. This aligns with the finding in the original paper, which means that my sample and calculations do not obviously bias any further analysis I do against this result.

Physical Regressions

The physical data allows me to do the original regressions intended in Gao and Van Biesebroeck (2014a) without the confounding missing prices. While they have made efforts to control for this issue, it is entirely plausible that X_{it} did not adequately control for missing prices, and the coefficients in these regressions may still be biased.

Table 2.11: Gao and van Biesebroeck Specification Using Physical Data

VARIABLES	(1) Log Coal	(2) Log Emp
Log Output	0.84***	0.70**
	(0.16)	(0.29)
Restruc	0.092	-0.22***
	(0.060)	(0.074)
Restruc x SOE	-0.0054	0.077
	(0.045)	(0.079)
Log Capacity		-0.53*
		(0.28)
Constant	3.46*	4.80***
	(1.79)	(0.50)
N	2,879	2,634
Plants	739	731

Notes: Output is instrumented using market revenues. Includes plant fixed effects, but not employment interacted with province-level ones since prices no longer confound the data.

In these new regressions, there is no significant result in either equation. The point estimate for the labor equation is even moderately positive. While these are not extremely precise 0's and we cannot categorically rule out the results from the financial regressions, this casts serious doubt on the effects of restructuring. Even according to the analysis of Gao and Van Biesebroeck (2014a), the physical regressions are the preferred specification.

Pricing Regressions

Table 2.12: Naive Pricing Regressions Using G+vB Independent Variables

VARIABLES	(1) Log Price	(2) Log Price	(3) Log Input Price	(4) Log Input Price
Log Output	1.18***	8.94	.496***	1.42***
	(0.19)	(6.11)	(.103)	(.38)
Restruc	-.40***	-1.89	-.158**	-.28***
	(0..36)	(1.42)	(.054)	(.11)
Restruc x SOE	.36***	1.39	.181***	.27***
	(0.087)	(0.99)	(.052)	(.09)
Log Capacity		-8.2		-1.34***
		(5.6)		(.35)
Constant	-12.8***	-.60	-6.87***	-.57
	(1.79)	(3.62)	(1.21)	(.69)
N	2,621	2,597	1,957	1,946
Plants	560	559	503	503

Notes: Output is instrumented using market revenues. Includes plant fixed effects, but not employment interacted with province-level ones since prices no longer confound the data.

These pricing regressions get at why the physical and revenue-based assessments of restructuring differ: input and (possibly separately or in turn) output prices are varying with the restructuring and treatment variables. This may be for several reasons: the restructuring policies themselves may have lead to these pricing changes, or separate deregulation in input markets merely make it look like the reforms were successful.

Planner Behavior Regressions

Table 2.13: Planner Wedge Regressions

VARIABLES	OLS	OLS	OLS	IV
Linear Marginal Cost	-.21***	-.55***	-.30*	-1.2**
	(.07)	(0.11)	(.17)	(.51)
MC x Restructuring			-.34*	
			(.18)	
Restructuring	-.02	.13***	.19***	.16***
	(.02)	(0.03)	(.04)	(.03)
Restruc x SOE	.05**	-.1**	-.10**	-.05
	(.03)	(0.04)	(.04)	(0.05)
Constant	.09***	.08***	.03	.21**
	(.02)	(.03)	(.04)	(.09)
Fixed Effects	None	Plant	Plant	Plant
N	2,090	2,090	2,090	1,498
Plants	526	526	526	428

Notes: Dependent variable is residual allocated production from planner behavioral model. Instruments include lagged marginal cost and investment behavior of largest firms in each province.

These regressions shed some light on why restructuring may not have resulted in large efficiency improvements: with plant-level fixed effects included, there is no evidence that, controlling for marginal costs, SOE plants experienced more favorable draws from regulators. If anything, 3 of the 4 specifications indicate that things may have gotten worse for them. Regressions of marginal cost on SOE and restructuring status suggest that SOE's actually improved their costs at a slower rate in response to restructuring. Thus, by available physical measures they did not become more efficient than private plants, and a better return on marginal costs may have actually hurt them. The point estimate on restructuring alone is positive, however, suggesting there was not overall cost savings in response.

Robustness Checks

Gao and Van Biesebroeck (2014a) do a large number of robustness checks to support their findings, and get consistent results across almost all of them. Given that this paper is focused on additional analysis, I have chosen only to run the physical counterparts for a select few.

As mentioned before, my results hold (as in, no significantly negative coefficients) using provincial output, lagged output, and twice-lagged output as instruments. Importantly for my analysis, they are also robust to only looking at plants that have full price data, which significantly decreases the available observations. OLS versions of the physical regressions also fail to replicate the financial efficiency findings, as does restricting the restructuring dummy to 2004 or later instead of 2002.

It is possible that a detailed event study or a similar analysis could reveal a common reason that invalidates both the original and my analyses. At any rate, the new regressions show that the case for restructuring leading to plant-level efficiency gains is ambiguous at best.

2.8 Analysis and Counterfactuals

2.8.1 Reallocation via Planner Behavior Model

As done in Eisenberg (2019), this counterfactual eliminates all variation in μ to simulate a planner that cares only about costs. Overall, the cost savings per unit across the plants included in the sample is about 2.8%. This is much lower than the accounting-cost based measures presented later in this section, which suggests unobserved cost shocks and heterogeneity may be playing a significant role in how planners behave in China.

2.8.2 Optimizing Firms

First, we can assess the changes in aggregate production that would happen if firms were simply allowed to produce according to their optimality conditions. In the year 2000, if every firm produced optimally, aggregate in-sample production would fall by 190 million megawatt-hours. Observed aggregate production was just under 300 million, so this represents a dramatic decline. 150 firms are included in the 2000 sample, and only 6 of them would shut down under optimal production. This suggests that pre-2002, most plants were operating basically sustainably, but overall the market was producing more than it would have collectively liked to.

This picture is somewhat different in 2005. In-sample production is now closer to 900 million (which could be due both to growth and to sampling patterns being more favorable), and optimal production would actually be about 160 million megawatt-hours **larger**. In addition, 54 plants of 276 would actually shut down if they were to produce optimally. This suggests that some serious misallocation is happening: even though almost 1/5 of the total stock of plants would shut down, aggregate production would still optimally be 150% larger. Cost-efficient firms are likely being constrained by pricing rules, while these same rules are not enough to fully prop up less efficient firms.

2007, the latest year in the data, looks similar to 2005. Optimal production is roughly 200 million mwh's larger, and 80 of 360 firms would shut down. So, 2005 does not appear to be any kind of outlier.

2.8.3 Social Planner Problems - Accounting Cost Minimization

Given my estimation and model, I get a firm-level measure of misallocation $\tilde{q}_i$ (or τ_i). To get an idea of how "far" this is from an optimal result, I compare it to a social planner's solution where the planner is free to move quantities around across plants. To start, I do this province by province, since for most of my sample that seems to be the appropriate market definition. It will also give me a more conservative estimate of misallocation (than a grid-based or national market) in a sense–the transmission infrastructure within each province is more robust than cross than across provinces. This should also give me an estimate that is free of province-level local protection.

The key to this social planner's solution is that I constrain the aggregate production to be the same as the observed total in the data. The primary motivation behind this is that electricity supply in China tends to be planned out across a year, and demand is extremely inelastic. Thus, the idea that there is a yearly "target" quantity that governing authorities are trying to reach is a reasonable approximation. Any temporary shocks that affected demand over this period will still be incorporated into this counterfactual, because they are reflected in the aggregate demand for the year of data.

The first social planner's problem can be written in the following way:

$$\min_{q_i'} \sum_i^N C_i(q_{opt,i} + q_i') \tag{2.9}$$

$$s.t. \sum_i^N (q_{opt,i} + q_i') = \sum_i^N (q_{opt,i} + \tilde{q}_i) = Q_i \tag{2.10}$$

Where C_i is each firm's cost function, and $q_{opt,i}$ is the firm's optimal quantity as determined by their profit maximization problem. This is equivalent to taking the most cost-efficient firm, raising their output until their marginal cost gets too high or their capacity is hit, and then moving down the efficiency order in succession. It is essentially the "sorting algorithm" from Asker et al. (2017).

Undertaking this problem for the year 2000 and summing across all sampled firms in all provinces gives a potential direct cost savings of $900 million dollars, on top of an original (fitted) total cost of $7 billion. While 15% is not a staggering number, the absolute amount of savings to get the exact same output is quite large. If unavoidable frictions account for even 80% of this value, regulations left at least a billion dollars on the table in 2000.

If one adds emissions savings, another $10 million (275 million upper bound) is gained. With the price distribution remaining identical, plants gain $270 million in profits. Sampled firms are roughly half of installed capacity for the year 2000, so if one assumes a constant rate of misallocation the savings could become 18 billion in direct savings, 20 million in emissions savings (55 million upper bound), and $540 million in profits. Since there is no easily testable direct assumption to make on the degree of misallocation for plants with missing data, one can consider the first set of numbers to be a lower bound.

In 2005, the savings are roughly 2 billion out of 24 billion in total costs, with sampled firms representing about 55% of installed capacity. An interesting wrinkle is that overall profits would actually **decrease** under this allocation (since prices are kept fixed). Emissions savings would net about $93 million ($240 million upper bound). While the change in profits is qualitatively different than the situation in 2000, the basic facts are similar: a large amount of money via both direct costs and emissions are not being realized, and the net sum of regulations (on input/output prices as well as production decisions) seems to be leading to large amounts of misallocation.

2007 continues the trend: \$3.4 billion in savings on top of \$40 billion in costs is possible, with another \$90 million (.24 billion upper bound estimate) from reduced emissions.

2.8.4 Accounting Cost Minimization Within Regional Grids

I also repeat this exercise using China's six regional grids, rather than provinces, as the reference market. This serves several possible functions: First, it provides an even more aggressive estimate of the savings that can be achieved under superior input allocations. To the extent that the Chinese government has aggressive plans to implement transmission networks (which they have had before and currently do), it is important to know what sort of improvements this can get us under existing infrastructure. Second, it can give us an idea of how prevalent cross-province local protectionism is compared to within-province issues. If massive savings are possible from redistributing inputs more easily within grids, it suggests that reducing the role of provincial authorities could be effective policy. Third, the grids became more relevant in the wake of restructuring, so grid-level analysis may provide some further insight into how restructuring affected input allocation.

This exercise nets a savings of \$1.35 billion in 2000 on top of an original \$7 billion in costs. The savings are especially pronounced this early in the sample, because there are multiple provinces with only one plant, which means there are no possible gains to be made by reallocating their inputs if things are restricted to the province level. Profits see a similar increase, from \$.27 billion to \$.38 billion. So, under existing output prices, there are dramatic profit gains to be made by reallocating inputs within the markets the Chinese government has designated for itself.

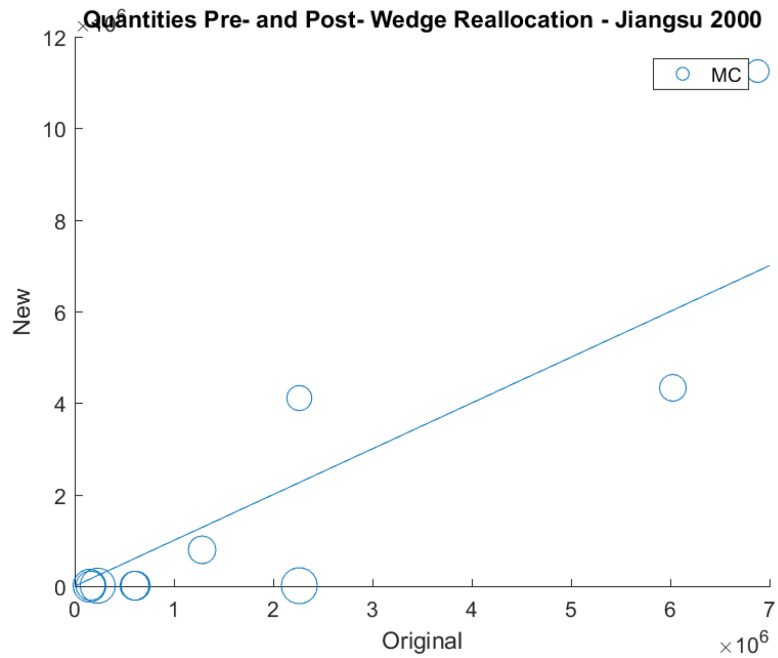
Importantly, emissions would actually **increase** under this allocation. This suggests that there are key input price differences across provinces that this allocation would take advantage of that trump whatever differences in heat rates there are. Within provinces, input prices are correlated enough that it is easier to make real savings from moving inputs to more physically efficient plants. There are no differences between the within-province and within-grid counterfactuals in terms of which plants are closed.

2005 sees a similar gain: cost savings go from about \$2.1 billion to \$2.5 billion. Profits, however, would decrease by almost \$ .3 billion. This is not necessarily shocking, since profits decreased even in the within-province reallocation scenario. The emissions increase seen in 2000 is now gone, and emissions damages would decrease by nearly \$70 million. 54 of 276 plants would be kept at 0 production in this scenario as well.

In 2007, recall that total in-sample costs were \$40 billion, and within-province reallocation would result in savings of \$3.4 billion. Within-grid reallocation increases this only modestly to \$3.8 billion. While this is only a 1% increase in savings, \$.4 billion is quite large and could easily outweigh the costs of more power lines. This increase is actually similar in magnitude to the increase in 2000, which suggests that they may be persistent differences across provinces that are being picked up by this calculation.

As before, all of these numbers can be in some since considered conservative, since my sample includes only about 50% of installed capacity. A more aggressive prediction would be to roughly double all of the numbers above.

Figure 2.2

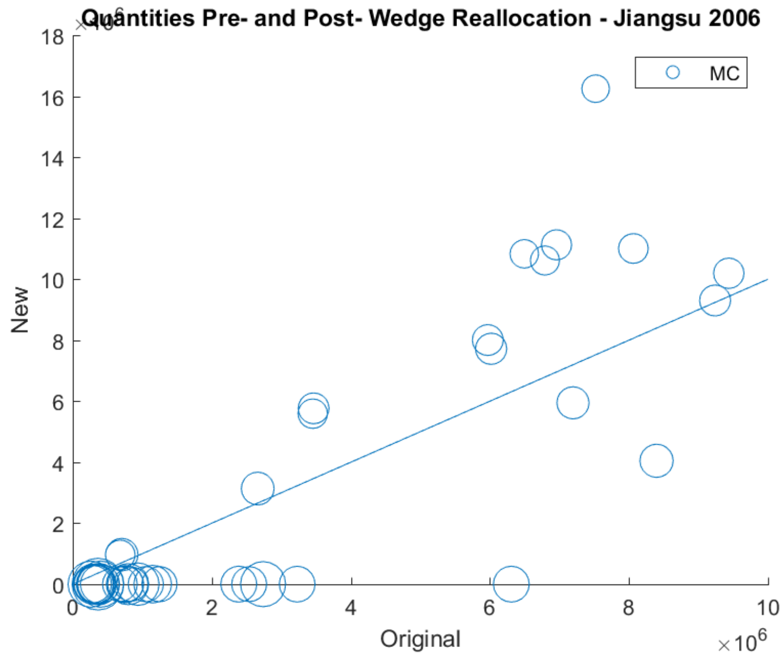


2.8.5 Jiangsu: A Case Study

It is clear at this point that there are large cost savings to be had at the national level via reallocating inputs, no matter the scope we choose. What is less clear is where exactly these savings are coming from: broadly speaking, we know inputs are moved from less efficient plants to more efficient plants, but by how much? How exactly are input prices and output prices playing in to all of this? To better answer these questions, it is useful to look at some firm-level changes rather than just aggregate measures from the counterfactuals.

I chose the Jiangsu province because it has a reasonable number of plants in all years of my sample, a fairly broad plant size distribution, and a mixture of private and Big 5 plants.

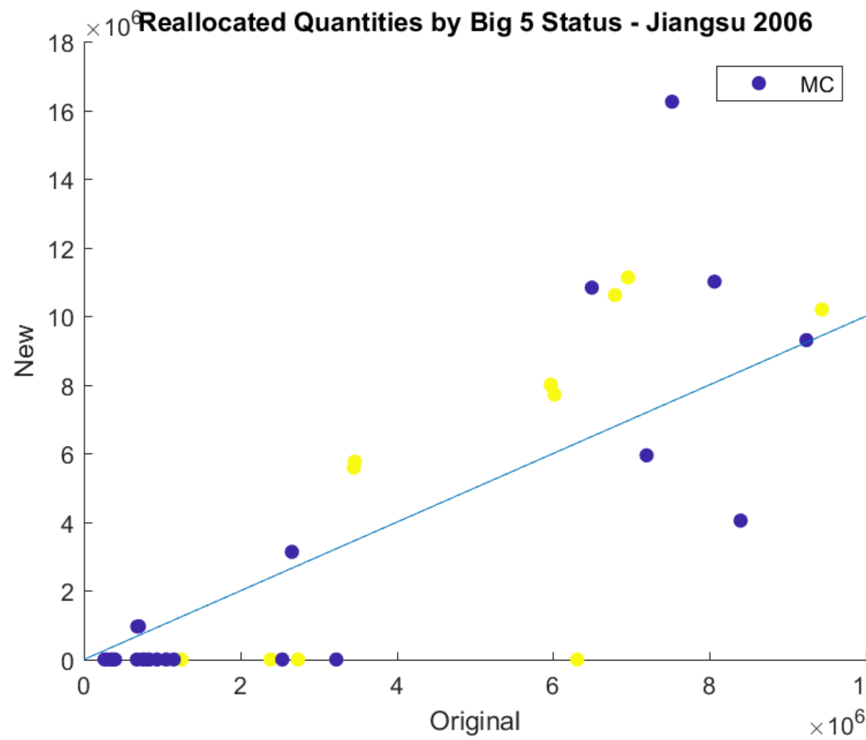
Figure 2.3



These pictures don't show identical trends necessarily, but the broad picture is true across both of them: generally speaking, both before and after restructuring, the cost-minimizing reallocation would involve nearly shutting down many smaller plants and moving their production to larger plants. Larger plants tend to have lower heat rates on average, and they incur additional savings due to the "threshold" effect in the cost function. The marginal cost increase from having a higher utilization is greatly diminished if a firm has larger capacity. There are also hard capacity constraints involved here: only so much can be physically reallocated to smaller plants, so even if they were more efficient it would not necessarily be feasible to shift away very much output from larger plants. This is part of why it is rarer to be below the reference line for very large plants.

This graph illustrates the difference between "Big 5" firms and others in the 2006 reallocation. While there is a mixture of quasi-private and and big 5 plants above and below the line along the size distribution, it is clear that many smaller private firms

Figure 2.4



would be nearly shut down, while only a few Big 5 plants would be. At any rate, a much higher percentage of Big 5 plants would survive.

This starts to reveal some of the underlying reasons why China has been reluctant to change its regulatory regime–to the extent that a competitive regime would reflect this planner's solution (not a guarantee, but plausible), incumbent Big 5 firms stand to benefit the most, while smaller private firms would lose out. The government has been trying to induce private entry into the market over the past couple of decades to meet capacity needs, and a market setup where they are trounced by incumbent SOEs would clearly deter this. Heavy price and quantity controls seem to achieve a measure of equality across plants even if this means that many are forced to operate mildly unprofitably.

2.8.6 Profit Maximizing Firms and Socially Planned Prices

Another set of counterfactuals can help shed light on the role output prices are playing in misallocation. Thus far we know that there are extremely large cost savings and (sometime) profit gains to be made by simply pursuing a cost-minimizing allocation across plants. This is obviously only one facet of the regulatory regime's behavior. There are many incentives to keep output prices low: namely that it helps facilitate lower retail prices for residential and industrial users, and minimizes losses for the government-controlled grid operators.

Using the same aggregate quantity constraint as before, we can run the following counterfactual: suppose that each plant is free to optimize, but takes prices. That is, τ is set to 0 for every plant, and they produce $q^*(p_i)$. Rather than maximizing over quantity distortions, the social planner now chooses prices for each plant that minimzes aggregate cost (since each firm is profit-maximizing, the aggregate cost-minimizing allocation will also maximize aggregate profits).

Mathematically, this problem can be represented the following way:

$$\min_{p'_i} \sum_i^N C_i(q^*(p'_i)) \tag{2.11}$$

$$s.t. \sum_i^N (q^*(p'_i)) = Q_i \tag{2.12}$$

Where p'_i is the new vector of prices, C_i represents cost functions, and q^* is the firm's optimal quantity based on their price (which will either be 0, satisfy their FOC, or be their capacity).

Note: national results for this are still forthcoming. My new cost specification has made this code take a while. My initial results suggested that overall massive price decreases OR increases were not necessary, but rather that modest shifts were needed to (yet again) favor the larger plants.

2.9 Conclusion

This paper provides a first step in documenting and diagnosing the misallocation in the Chinese coal power industry. This manifests in several ways, with firms producing far from their optimums based on first-order conditions, to prices being set in such a way that aggregate production would be drastically different if they were allowed to. Optimally producing firms, by my estimates, would cause output to fall by over 50% in 2000, and rise by roughly 20% in 2005 keeping prices fixed. That many plants would like to optimally shut down in both years just accentuates how inefficient cross-plant allocation of resources were at this time. It seems clear that whatever regulatory regime was in place, both before and after restructuring, was not geared toward getting plants to produce at efficient levels, even if these estimates are also picking up many unobserved frictions.

These plant-level distortions result in aggregate inefficiency: both before and after restructuring, if a social planner were allowed to reshuffle inputs to meet the same production quota while minimizing costs, there would be national cost savings of at least 9% (and sometimes up to 15%). If a social planner is allowed to reshuffle within regional grids, these figures jump to nearly 20%. This signals the urgent need for improved transmission infrastructure in China, and sheds light on the possible role that province-level local protectionism may be playing in this market.

These results are starkly contradicted when a structural model of planner behavior that includes unobserved cost shocks is applied. Using this approach, aggregate gains from efficient reallocation fall closer to 3%. Given the conservative nature of this approach, it is likely that the actual amount falls somewhere between my two estimates.

The environmental implications of these findings are small compared to the absolute financial figures coming in to play, but savings from emissions reach almost $1 billion under some counterfactual scenarios, and up to $2.4 billion under some stronger assumptions when using only accounting costs. If we extrapolate these to the capacity growth that has happened since 2007, it is possible that these numbers are now dramatically larger.

While there are competing explanations for many of the findings in this paper (frictions, lack of enforcement, local protectionism, measurement error), it establishes a firm baseline result that the 2002 restructuring did not cause any fundamental improvements in how resources are allocated across power-plants, even if there was a general move toward lower heat rates at the time. This is in line with the many qualitative papers that suggest these reforms did not quite "take", and that many institutional changes need to be made before China would have a chance of benefitting from a Western-style deregulation [7]. So, while current policies appear to be leading to large-scale misallocation and economic losses, China's skepticism in this regard is likely warranted.

[7] Especially given that it is unclear how much Western economies benefitted from these changes.

CHAPTER 3

HETEROGENEOUS TECHNOLOGIES, PRODUCTIVITY AND THE STATE SECTOR IN CHINA

3.1 Introduction

The manufacturing sector in China has been the subject of extensive study in recent years. It has seen massive productivity growth, large amounts of private entry, and a comparably massive of exit and privatization of its state-owned firms since the 1990s (see Brandt et al. (2012) and Song et al. (2011), among others). However, given that the Chinese manufacturing census data has only recently become widely available, many key questions regarding this process are still not well understood.

Broadly speaking, it is understood that both state-owned and private firms in China saw high amounts of total factor productivity growth from 1998 to 2007. It is also generally understood that state-owned firms (SOEs) have historically had generally lower TFP than private firms, but in recent years have been growing at roughly the same rate, or even slightly faster (though they still remain "behind" private firms). While the patterns of state ownership across different industries in China are complex and varied, it is generally the case that state-owned firms have more favorable access to capital markets. As per Song et al. (2011), this has far-reaching welfare implications and leads to substantial systematic differences between the two types of firms. In that paper, some of these differences are imposed as an assumption, while this paper seeks to recover their magnitudes using micro estimation techniques from the industrial organization literature.

This paper posits that having differential access to capital markets would influence the technology choice of SOEs, and their Cobb-Douglas production functions would be measurably different than those of private firms. There are arguments to be made that SOEs should be either more or less capital intensive than private firms. To argue the former, one could say that when an SOE makes its initial choice of technology, it does so knowing it has favorable access to capital markets and will for years to come. Assuming that "technology" as represented by the production function takes many years to change, or is subject to heavy adjustment costs, it may be optimal for the firm to invest in a technology where the more freely available capital has a higher return (and thus a higher coefficient).

If one takes the view that technology is more flexible, or takes more seriously the view that a CD production function is a reduced form aggregate of several different technologies, many of which can be changed more freely, then SOEs being less capital intensive makes more sense. Since capital is more freely available to them, they will "overinvest" as compared to private firms, and, assuming that there is eventually diminishing returns, they will have lower marginal products with respect to capital. This would manifest as a lower capital coefficient in the production function.

The results in this paper support the notion that favorable access to capital markets leads to measurably different technological choices for SOEs, according to the coefficients recovered from their Cobb-Douglas production functions. Specifically, they are estimated to be substantially less capital intensive. In turn, this has implications for many key results on productivity. For example, the relative drain that misallocation–the notion that less productive firms have unduly high market shares–has on aggregate SOE productivity decreases by almost 10%.

While the number of firms in the state sector has fallen drastically in recent years, the remaining SOEs tend to be very large and powerful. They are often seen as either market leaders or prominent public companies in China (see Hsieh and Song (2015)). In 2007, the year when SOEs represented the smallest percentage of firms in the economy in our sample, they still maintained a 30% revenue share. So, while SOEs are currently a shrinking population, they still represent an incredibly large fraction of output.

In moving from more macro-oriented models to micro estimation methods to address these question, this paper loses some prescriptive power. Essentially, since the goal of the paper is to capture and document certain facts under a weaker set of assumptions than they are usually studied, it is more difficult to compare the outcomes that the methods generate without these stronger assumptions. As a result, it is necessary to carefully examine which results are numerically comparable, and what signs and magnitudes are meaningful. The decompositions from Pakes and Olley (1996) and Melitz and Polanec (2013) are used repeatedly on different subsamples to aid in this.

3.2 Literature Review

3.2.1 (Mis)Allocation and State Ownership

Two of the closest papers to this one in the literature are Hsieh and Klenow (2009) and Hsieh and Song (2015). The former paper backs out firm-level and aggregate TFP from a series of modeling and optimization assumptions, including first-order conditions for labor and capital and CES aggregators across firms and industries. With this information in hand, the authors use comparable information from the US to establish a benchmark amount of "misallocation" - which is defined as deviations from the optimal allocation of resources as defined by first order conditions and firm-specific TFP levels. Firms are assumed to have constant returns to scale production functions, and in a robustness check they allow each firm to have their own labor and capital coefficients, as defined by their labor revenue share of output (with constant returns to scale still imposed).

Similarly Hsieh and Song (2015) use this framework to examine many of the same questions that this paper addresses. Namely, they use a model of optimizing firms, input and output distortions, and CES aggregators to back out analytical solutions for TFP, labor productivity, and capital productivity. They then analyze these patterns across state ownership status over time. The authors find that state-owned firms generally have lower capital productivity than private firms, faster TFP growth, and comparable labor productivity. They then run a series of counterfactuals to estimate the extent of misallocation due to the state sector. However, these different productivity terms still only incorporate one source of technological heterogeneity: the TFP residual.

This paper can partially be seen as a bridging of the Hsieh and Klenow (2009) and Hsieh and Song (2015) framework with the micro productivity estimation literature. It extends the questions of Hsieh and Song (2015) to see if their conclusions are robust to a more relaxed set of firm-level assumptions (and thus more complex estimation routines). In addition, if their results translate to this new framework, this paper helps to refine the magnitudes of the central objects of concern, like labor/capital productivity gaps and measures of misallocation. It additionally explores how sensitive these results are to various forms of heterogeneity in production technologies that are captured via micro estimation.

There are several other papers that analyze productivity, allocation, and state ownership in China, or use similar macro frameworks in other countries to explore the issue more generally, such as Brandt et al. (2013) and Restuccia and Rogerson (2008).

3.2.2 Structural Productivity Estimation

This paper borrows heavily from the proxy variable productivity estimation literature, famously pioneered by Pakes and Olley (1996). Across the different specifications, the estimation borrows elements from Olley-Pakes, Levinsohn and Petrin (2003), and Ackerberg et al. (2005). The unifying theme of these papers is that the productivity residual is endogenous in a regression that estimates a production function (because the firm observes it at the time they make their output decision), so additional structural assumptions and estimation routines are necessary to deal with this so-called "simultaneity bias".

Another advantage of these methods, and one that helps to distinguish the paper from the macro models that do not exploit these estimation routines, is that they do not require labor and capital to be flexible inputs. That is, the results do not rely on any first-order conditions that reflect same-period decisions, and the recovered production function parameters and TFP residuals could be reflective of firms who are behaving optimally according to a dynamic programming problem.

This paper also connects to the recent literature on non-Hicks neutral productivity shocks, like Balat and Sasaki (2014) and Zhang (2015). Both of these papers have similar aims: to estimate production functions where firms can be heterogeneous beyond a simple TFP residual. In a future iteration of this paper. These papers relate to a more general critique raised recently by Gandhi et al. (2013), which argues that a large percentage of productivity residual heterogeneity may be due to misspecification and non-identification of many of the production functions assumed in the literature [1].

[1] For now, this paper uses value-added production functions and is subject to the non-identification critique presented in Gandhi et al. (2013), though I am working on addressing this with alternative specifications.

3.2.3 General Productivity Papers in China

This paper also borrows from and relates to several more general productivity papers that focus on China. Both Brandt et al. (2017) and Brandt et al. (2014) have provided publicly available algorithms for working with the Chinese manufacturing census data, which this paper heavily exploits. The "Data" section will explain more in-depth how these papers are used.

Gao and Van Biesebroeck (2014b) looks at the Chinese electricity industry and creates difference-in-difference estimators based on changes on deregulation across different regions. While this paper takes a broader look at the Chinese economy, its examination of the margins of privatization and nationalization overlaps with their smaller, more specific analysis. Brandt et al. (2012) looks at productivity shifts in China related to its joining the WTO. While this iteration of this paper does not evaluate any specific policies other than state ownership, this would be a good direction for further research one the main analysis here is refined.

3.3 Data

The data in this paper come from the manufacturing census conducted by China's National Bureau of Statistics. The census contains yearly reports from 1998 to 2007, recording financial at production at the firm level across China. The sample contains all firms with sales above 5 million RMB. While this unfortunately excludes 80% of industrial firms in China, it includes over 90% of industrial output and over 70% of employment. So, while many small firms are excluded the surveys represent the majority of economic activity in China's manufacturing sector [2].

The major observed variables of interest at the firm-level for this analysis are as follows: revenue, wage bill, employee benefits, fixed assets, intermediate inputs, financial expenses, a firm's unique identification number, various capital ownership variables, a firm's "registration type", and basic geographic identifying information.

There are several major cleaning considerations in working with this dataset. First, the micro-level observations are not matched cleanly from year to year. Second, most of the variables of interest are in nominal terms, or are at odds with comparable results in aggregate datasets, and need to be addressed. Third, the categorization of which industry each firm belongs to changes during the sample period.

3.3.1 Matching Panel Observations

Structural productivity estimators usually exploit the panel dimension of a dataset. In this case, there are unique identifiers that are meant to track the same firm across different years, but they are not always reliable. Some of this has to do with changes in the structure and ownership of firms, and some of this has to do with measurement error. As a result, it is necessary to try and match firms over time by more than just their unique identifiers (FRDM), though the majority of matches are achieved this way. According to Brandt et al. (2012), roughly 95.9% of matches can be achieved using the firm ID.

In general, firms are matched using the following algorithm from Brandt et al. (2014): Firms are initially matched based on their numerical identifier across each set of two consecutive years. For firms where this identifier is a duplicate value within the

[2]These figures are from Brandt et al. (2012), and based on the full census that is performed every 5 years (2004 for the most recent iteration in the data) where all firms are included

same year (roughly 10 to 30 observations per year), the firm's name is also used.

When the FRDM and name matching fails, observations are matched via different combinations of firm name, the firm's "legal representative" name, phone number, address, product name, "geographic code", industry code, and founding year. While the sample that is generated does not match up 100% with Brandt et al. (2014), the patterns and sizes in each year and firm type are in line with their findings.

To address the fact that some firms may fall out of the sample briefly due to measurement error (or falling below the 5 million RMB threshold), the unmatched firms from each two-year sequence are also compared to observations two years away in the data. Then, the full set of 3 year matches are reconciled and stacked into a full 11 year panel.

3.3.2 Concordances

An important aspect of this dataset is the categorization of firms into different industries. This is done via the China Industrial Code, which is traditionally measured at the two or four digit level. While the four-digit level is the most specific, most four-digit industries do not have sufficient observations for the type of inference done in this paper. Hence, the analysis will suffer from some level of aggregation bias, though this is hopefully mitigated by differentiating at the 2-digit level.

As an example, the two digit code 34 refers to "Fabricated Metal Products", which incorporates a number of more specific four digit industries. Industry 3421 is "Cutting Tool Manufacturing", and Industry 3433 is "Metal Packaging Containers Manufacturing" [3].

[3]These names are translated from the original Chinese.

Firms that operate in different industries are seldom directly compared, and we must allow for different production functions for each industry. A major issue in this dataset is that the definition of industry concordances (CIC) changed officially between 2002 and 2003. I borrow, again from Brandt et al. (2012) and Brandt et al. (2014), a mapping from pre- and post-2002 CIC's to a unified set that works across all years. For now, since my analysis takes place at the 2-digit CIC level (as opposed to the more refined 4-digit specification), this is mitigated since most industry definitions did not change across the 2-digit specification.

3.3.3 Deflators

Several key variables, such as output and the capital stock, are listed in nominal terms. Since output prices are not observed, it is necessary to create industry-specific deflators to get the closest approximation the real output possible. Calculation of firms' real capital stock and deflators for output, capital, and intermediate inputs are done according to the algorithms and deflators shared in Brandt et al. (2012) and Brandt et al. (2014). The dependent variable for most of the analysis, value added (denoted in most equations by y_{it} is calculated via the "double deflation method". Deflators are calculated for output and input separately, and then value added is real output minus real inputs.

In order to get the most specific possible deflators, I used price deflators calculated at the 4-digit CIC level whenever possible. If these deflators were missing I opted for one calculated at the 2-digit level instead [4].

[4]Some industries with two-digit codes 40 and above are still missing deflators, though they account for a relatively small share of output.

3.3.4 Labor

The labor variable is constructed by taking the sum of a firm's total wage bill and benefits paid to employees. In Brandt et al. (2012), additional insurance variables are added, but this paper excludes them as they do not consistently appear in the sample. Labor is then inflated by a constant factor so as to match the share of value added published in aggregate datasets such as the China Statistical Yearbook [5].

3.3.5 Summary Statistics

Table 3.1 below provides some summary statistics on basic variables of interest for the analysis.

Table 3.1: Summary Statistics

Year	Output	Employment	Fixed Assets	Number of Firms
1998	6.0	53.6	5.1	157,025
1999	6.7	50.1	5.5	154,022
2000	7.8	47.9	5.9	148,815
2001	8.9	47.2	6.1	157,647
2002	10.7	48.3	6.4	168,990
2003	13.8	50.5	6.9	178,715
2004	18.2	59.1	8.1	265,006
2005	21.9	61.9	9.0	258,333
2006	27.0	66.0	10.3	283,045
2007	31.4	64.5	11.0	260,393

Output (deflated sales) and Fixed Assets are in trillions of RMBs, while employment is in millions of workers. Differences from Table I in Brandt et al. (2012) reflect the fact that theirs includes some firms outside of the manufacturing sector. Output and capital are in real terms.

A few basic trends are clear: there is a lot of net entry over time, and the use of both

[5] In Brandt et al. (2012), this is roughly $\frac{.55}{.32}$. This paper's labor shares do not quite match this, but I borrow the same inflation factor to get a first-order approximation of the same labor variable.

labor and capital inputs is increasing. Trends from, for example, 2007, however, suggest productivity could also be growing. Output appears to have increased despite a decrease in employment and the number of firms.

3.3.6 Identifying SOEs

Once firms are identified through time and outputs and inputs are properly deflated, a key next step is to identify which firms are owned by the state. This is more or less observed in the data: there is a variable called "registration type," and several of the values it takes on explicitly correspond to state ownership.

Using only this variable, however, understates the presence of SOEs [6]. It is also necessary to see who own the largest shares of capital in each firm. This can be done via six key capital variables. The data tells us the total amount of capital a firm has received, and the various amounts that is has gotten from the government, from collectives, from "legal entitities", from individuals, and from foreign nations.

In addition to firms who are explicitly listed as SOEs, a firm is an SOE if its capital received from the government is both larger than what it received from legal entities and individuals, and it is also listed as a "stock limited company" or an "other limited liability company" [7]. This essentially identifies firms where the government is the primary shareholder and thus controls the company, even if they are not formally registered as a government-run company.

This methodology results in the following state ownership patterns:

[6]See Hsieh and Song (2015) for details.

[7]I thank Yifan Zhang for the code to apply this method.

Table 3.2: SOEs Over Time

Year	SOEs	Private	SOE Proportion	SOE Revenue	Priv. Revenue	Revenue Proportion
1998	60319	96706	0.384	2.769	3.261	0.459
1999	56466	97556	0.367	3.001	3.736	0.445
2000	49504	99311	0.333	3.386	4.420	0.434
2001	44273	113374	0.281	3.712	5.198	0.417
2002	41002	127988	0.243	4.231	6.507	0.394
2003	37198	141517	0.208	5.085	8.717	0.368
2004	46641	218365	0.176	6.037	12.18	0.331
2005	39579	218754	0.153	7.232	14.63	0.331
2006	38462	244583	0.136	8.302	18.70	0.307
2007	31631	228762	0.121	9.166	22.27	0.292

Revenue is in trillions of RMBs, and in real terms.

While the number of SOEs in the economy is rapidly shrinking during the sample, they remain a major force in the economy for its entire duration. Notably, SOE revenue grows steadily each year even as the number of firms shrink, and they retain a 30% overall revenue share despite the mass exit that is taking place. Thus, the SOEs that remain are incredibly large, much moreso on average than the private firms in the economy.

3.4 Model

3.4.1 Baseline

The "baseline" specification is a straightforward application of the Ackerberg et al. (2005) method. It works according to the following model:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \varepsilon_{it} \tag{3.1}$$

This represents a standard Cobb-Douglas model: y_{it} is log of firm-level value added (deflated revenue minus deflated intermediate inputs), k_{it} is log real capital stock, and l_{it}

is the log of the firm's wage bill [8].

ω_{it} represents the firm's known total factor productivity. That is, it is observed to the firm at the time it makes its input choices, but not observed to the econometrician. This is a key component of the model, as it implies that input choices may be correlated with the unobserved error term. ε_{it} is an unexpected productivity innovation, unobserved to both the firm and the econometrician (and hence not as much of a problem for estimation).

As is well established in the structural productivity literature, this model cannot be estimated using standard OLS or fixed effects regressions. The ACF framework imposes a few structural assumptions so as to facilitate estimating the Cobb-Douglas coefficients and TFP residuals for all firms.

The first set of assumptions in the ACF framework are on the timing of input choices, and their relationship to TFP:

1. At time t, when the firm makes its output decision, its capital stock k_{it} and labor supply l_{it} are already fixed and known to the firm [9].
2. There is a flexible input, such as intermediate inputs (denoted as m_{it}), which the firm makes its decision on in time t.
3. The level of this flexible input is monotonic in TFP, like in Levinsohn and Petrin (2003). So, if we write $m_{it} = f_t(\omega_{it}, k_{it}, l_{it})$, then this assumption allows us to assert the following equation: $\omega_{it} = f_t^{-1}(m_{it}, k_{it}, l_{it})$.

[8] See the Data section and Brandt et al. (2012) for details

[9] The original paper has labor being "quasi-fixed" as compared to capital, and hence the decision is made slightly later, but most of the methodology is consistent with either set of assumptions. The only difference this makes is in the later stage of estimation we can treat same-period labor as pre-determined, while under other assumptions we would have to instrument using lags.

Given these three assumptions, we can rewrite our model the following way:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + f_t^{-1}(m_{it}, k_{it}, l_{it}) + \varepsilon_{it} \tag{3.2}$$

Or:

$$y_{it} = \Phi_t(l_{it}, k_{it}, m_{it}) + \varepsilon_{it} \tag{3.3}$$

Where Φ_t is some unknown function of the three inputs.

This allows us to do the first stage of the ACF estimator. Essentially, since $E[\varepsilon_{it}] = 0$ and it is exogenous, we can estimate Φ_t using standard non-parametric methods. In this case, I use a second-order polynomial sieve regression with a full set of interaction terms [10].

From here, we move to the second stage, which requires an additional assumption that ω_{it} moves according to a first-order Markov process. That is:

$$p(\omega_{it+1}|I_{it}) = p(\omega_{it+1}|\omega_{it}) \tag{3.4}$$

We can thus decompose same-period productivity into its expectation and an unexpected innovation term, ξ_{it}:

$$\omega_{it} = E[\omega_{it}|\omega_{it-1}] + \xi_{it} \tag{3.5}$$

[10] I plan to optimize the degree of the polynomial using cross-validation in a future iteration of this paper

So, we should expect ξ_{it} to be mean 0, and mean-independent of anything that is fixed at time t. Given our earlier assumptions, this means that $E[\xi_{it}|k_{it}] = 0$ and $E[\xi_{it}|l_{it}] = 0$. These moment conditions allow us to recover the coefficients of the production function in the second stage. Consider a candidate set of coefficients, β_k and β_l. For a given candidate pair, they imply a set of values for ω given our first-stage estimates:

$$\omega_{it}(\beta_k, \beta_l) = \hat{\Phi}_{it} - \beta_k k_{it} - \beta_l l_{it} \tag{3.6}$$

Then, a non-parametric regression of the estimated ω_{it} on ω_{it-1} should recover estimates of ξ_{it} [11]. We can then use these estimates combined with our two moment conditions above to create sample moments. The optimal estimates of our coefficients are those that minimize these sample moment conditions. With coefficient estimates in hand, we can then back out the corresponding ω_{it} values according to the above formula.

3.4.2 SOE Specification

The second specification, or "SOE Specification", allows for coefficients to differ across firm ownership categories. The basic model is as follows:

$$y_{it} = (\beta_l + \gamma_l SOE_{it})l_{it} + (\beta_k + \gamma_k SOE_{it})k_{it} + \omega_{it} + \varepsilon_{it} \tag{3.7}$$

Where SOE_{it} is an indicator variable for state ownership. This specification allows for heterogeneity across the two major different types of firms, while exploiting all

[11] Right now, this is done via a third-degree polynomial. The results do not appear to be sensitive to switching to either second- or fourth-degree polynomials.

within-industry information to get coefficient estimates. All of the timing and Markov assumptions remain the same for this specification.

This modification to the coefficients necessitates changes in both stages of estimation. When isolating the exogenous error term from the firm's initial policy function, it now takes the following form:

$$y_{it} = \Phi_t(l_{it}, k_{it}, m_{it}, SOE_{it}) + \varepsilon_{it} \tag{3.8}$$

SOEs and private firms are now allowed to have different policy functions in the non-parametric first stage, which is necessitated by this specification.

In the second stage, we now need to search over candidate β and γ values:

$$\omega_{it}(\beta_k, \beta_l, \gamma_k, \gamma_l) = \hat{\Phi}_{it} - \beta_k k_{it} - \beta_l l_{it} - \gamma_k SOE_{it} k_{it} - \gamma_l SOE_{it} l_{it} \tag{3.9}$$

We also need two additional moment conditions. If a firm's capital and labor stock are fixed at the beginning of period t, it seems fair to also say that their SOE status is as well. The assumption would be that whatever adjustment costs cause their capital stock to be fixed would also affect a large-scale transfer of ownership. Under this assumption, we can assert that $E[\xi_{it}|k_{it}SOE_{it}] = 0$ and $E[\xi_{it}|l_{it}SOE_{it}] = 0$. This gives us the four moment conditions we need to identify the four parameters of interest.

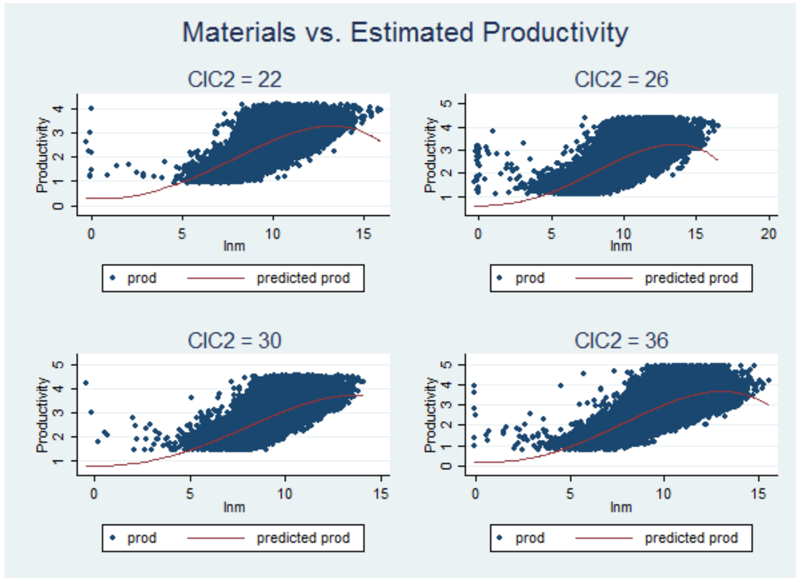
3.5 Results

3.5.1 Checking Assumptions and Cleaning

Before checking any assumptions or deriving any results, I cleaned the estimated sample. First, I dropped any firms for which log value added was missing (ie, any firms missing revenue, materials, or with a negative level of real value added). I then calculated each observation's productivity percentile at the 2-digit industry level. If a firm, in any year, had an observation that fell in either the upper or lower 1 % of their industry's distribution in either the baseline or SOE specification, I dropped them from the sample. While this gets rid of a substantial number of firms, and arguably those who could most affect the results, this ensures that the results are not due to any outliers facing exceptional circumstances. Post-cleaning, there are still well over 1.6 million observations in total.

There is an unfortunate side effect of these steps of cleaning: they disproportionately affect SOEs. In the early years of the sample, after all productivity estimation and cleaning has been accomplished, almost half of SOE observations are dropped. Additionally, earlier years are disproportionately affected so the mass SOE exit rate in the raw data is reduced for this sample. While comparable numbers of private firms are dropped, since they occur in much larger numbers a much higher percentage of them are retained. While it's unfortunate that the variable of interest is so heavily influenced by data cleaning, the hope is that we are retaining the middle of the SOE productivity distribution, and our results would be robust to over- and under-inclusion of outlying observations. In a future iteration of this paper, I plan to experiment with alternative sample selection regimes.

Figure 3.1



The first thing to check is whether the (testable) assumptions underlying Ackerberg et al. (2005) hold in our data. Given measures of raw materials and productivity, we can check (at the industry level, since each industry has a different policy function as per our estimation) whether the relationship is actually monotonic.

The figure below plots the relationship between materials and estimated productivity for four separate sample industries, along with fractional polynomial curves of best fit. We can see that, while it is generally a noisy relationship, and there are some slight nonmonotonicities at the tails, the best fitting function is monotonically increasing over the majority of the sample. Thus, while the relationship is not perfect, it appears that the simplifying assumption is generally satisfied.

3.5.2 Coefficients

There are two sets of key objects that each specification will produce: TFP residuals and production function coefficients. The coefficients are worth examining first because they provide a good gauge as to whether our estimation techniques gave us a reasonable answer. Also, if no differences are found across production functions, then re-examining the TFP results may not be necessary. Generally speaking, the production functions should be close to constant returns to scale, and there should be a reasonable amount of heterogeneity across 2-digit industries. The initial results are below:

Table 3.3: Baseline Coefficient Estimates

CIC2	Capital	Labor	RTS	N
13	0.235	0.511	0.747	96184
14	0.312	0.554	0.866	38612
15	0.287	0.601	0.889	25397
16	0.585	0.641	1.226	2057
17	0.243	0.456	0.699	138826
18	0.212	0.597	0.808	78752
19	0.240	0.523	0.763	39454
20	0.131	0.642	0.773	36098
21	0.165	0.639	0.803	19500
22	0.224	0.567	0.791	49981
23	0.332	0.656	0.988	35425
24	0.182	0.598	0.780	22445
25	0.423	0.382	0.806	2892
26	0.278	0.474	0.752	117023
27	0.313	0.531	0.843	32750
28	0.299	0.520	0.819	8239
29	0.293	0.464	0.756	19519
30	0.246	0.518	0.765	78403
31	0.216	0.527	0.743	138477
32	0.374	0.463	0.837	29645
33	0.346	0.378	0.723	20271
34	0.398	0.335	0.733	70344
35	0.223	0.528	0.751	124717
36	0.134	0.624	0.758	66897
37	0.254	0.657	0.911	79551
39	0.331	0.501	0.832	93766
40	0.274	0.637	0.911	54120
41	0.157	0.675	0.832	24449
42	0.229	0.499	0.729	33092
43	0.321	0.417	0.739	1661
44	0.451	0.621	1.071	42512
45	0.385	0.372	0.756	3047
46	0.448	0.526	0.974	19973

Results are from standard ACF specification with no accounting for SOE status.

There is a reasonable amount of heterogeneity across industries, ranging from a minimum returns to scale of .69 to a maximum of 1.22. The modal value of returns to scale appears to be around .75–this is low, and suggests there is likely a data issue with either capital or labor, especially given that it appears to happen across many industries. Comparable OLS estimates showed similar patterns: the returns to scale estimates generally fell within .08 of each other, though with considerable differences between labor and capital coefficients, and the ACF results were not uniformly higher or lower. This suggests that either capital or labor is probably under- or over-inflated somehow, thus depressing the returns to scale estimates. The underlying issue appears to affect all industries in the same direction. So, while the absolute magnitudes of any coefficient and TFP patterns we uncover may be subject to change because of this, their signs and relative magnitudes should stay relatively intact. In almost all industries, the labor coefficient is substantially higher than the capital coefficient, which is in keeping with the literature on value-added production functions.

3.5.3 SOE Coefficients

Since the baseline coefficients look reasonable, we can analyze how the SOE specification differs from them. Below is a table of the estimated capital coefficients for private firms and SOEs (and the associated difference between them). I have also included the number of SOEs and the number of firms in each 2-digit industry to show that the results are not just driven by a few industries that happen to be mostly state-owned.

Table 3.4: SOE Capital Coefficient Estimates

CIC2	$\beta_k^{Private}$	β_k^{SOE}	γ_k	SOE	N
13	0.318	0.112	-0.206	21511	96184
14	0.375	0.184	-0.191	9938	38612
15	0.370	0.152	-0.218	7901	25397
16	0.320	0.620	0.301	1748	2057
17	0.281	0.120	-0.161	12425	138826
18	0.220	0.168	-0.0525	3658	78752
19	0.268	0.0744	-0.193	1356	39454
20	0.208	-0.0148	-0.223	3686	36098
21	0.194	-0.0362	-0.231	1214	19500
22	0.270	0.152	-0.118	5607	49981
23	0.387	0.202	-0.185	13132	35425
24	0.207	-0.0285	-0.236	950	22445
25	0.415	0.430	0.0150	755	2892
26	0.340	0.191	-0.149	22060	117023
27	0.361	0.236	-0.125	10795	32750
28	0.351	0.216	-0.135	1049	8239
29	0.339	0.140	-0.198	2292	19519
30	0.285	0.106	-0.179	6075	78403
31	0.261	0.177	-0.0841	25496	138477
32	0.395	0.280	-0.115	4295	29645
33	0.371	0.240	-0.131	3121	20271
34	0.422	0.262	-0.160	6632	70344
35	0.301	0.0627	-0.239	18271	124717
36	0.251	0.0297	-0.221	14834	66897
37	0.326	0.128	-0.198	18948	79551
39	0.361	0.240	-0.121	11780	93766
40	0.297	0.162	-0.135	7407	54120
41	0.225	0.0288	-0.196	4606	24449
42	0.251	0.0738	-0.177	1534	33092
43	0.319	0.184	-0.135	136	1661
44	0.518	0.416	-0.102	34564	42512
45	0.410	0.354	-0.0551	1958	3047
46	0.462	0.398	-0.0634	17818	19973

Results are from ACF specification that allows for differing capital and labor coefficients by SOE status.

The results in this table are fairly stark. For all but two very small industries, γ_k is negative. For some industries, this leads to fairly unrealistic results: SOEs are estimated to have a negative capital coefficient in industries 20, 21, and 24. In industries 35, 36, and 42, the coefficient is very close to 0. Some of this is to be expected: not all industry-year combinations will have very many SOE observations, so our estimates will become very noisy for the parameters that are estimated almost only from SOE firms.

An encouraging sign is that the averages of the two coefficients (weighted by the relative number of firms) comes out somewhat close to the baseline estimate. For example, this comes out to about .27 in industry 13 compared to a baseline estimate of .24, or in industry 34 they average to about .41 with a baseline estimate of .398.

These results only capture the change in capital coefficients. On their own, they suggest that SOEs are on the whole have significantly lower capital productivity (as opposed to TFP and labor productivity, in this case). To put it another way, it appears that SOEs tend to adopt more labor-intensive technologies. Without the corresponding changes in labor coefficients, we cannot fully make this claim. To confirm this, we would need SOEs to also have higher (or similar) labor coefficients as in the baseline specification, and for estimates of returns to scale to not be substantially altered.

The following page showcases the labor estimates:

Table 3.5: SOE Labor Coefficient Estimates

CIC2	$\beta_l^{Private}$	β_l^{SOE}	γ_l	SOE	N
13	0.446	0.606	0.160	21511	96184
14	0.497	0.654	0.156	9938	38612
15	0.524	0.732	0.208	7901	25397
16	0.922	0.609	-0.313	1748	2057
17	0.438	0.585	0.147	12425	138826
18	0.589	0.633	0.0437	3658	78752
19	0.497	0.675	0.178	1356	39454
20	0.583	0.752	0.169	3686	36098
21	0.607	0.802	0.195	1214	19500
22	0.551	0.626	0.0756	5607	49981
23	0.599	0.727	0.129	13132	35425
24	0.575	0.754	0.179	950	22445
25	0.372	0.374	0.00127	755	2892
26	0.455	0.574	0.119	22060	117023
27	0.504	0.612	0.108	10795	32750
28	0.519	0.621	0.102	1049	8239
29	0.442	0.615	0.173	2292	19519
30	0.487	0.645	0.158	6075	78403
31	0.513	0.553	0.0394	25496	138477
32	0.442	0.570	0.128	4295	29645
33	0.358	0.501	0.143	3121	20271
34	0.312	0.476	0.164	6632	70344
35	0.501	0.697	0.196	18271	124717
36	0.576	0.743	0.166	14834	66897
37	0.624	0.785	0.161	18948	79551
39	0.480	0.595	0.115	11780	93766
40	0.618	0.744	0.126	7407	54120
41	0.632	0.795	0.163	4606	24449
42	0.484	0.647	0.163	1534	33092
43	0.407	0.573	0.166	136	1661
44	0.543	0.661	0.118	34564	42512
45	0.450	0.451	0.000530	1958	3047
46	0.593	0.590	-0.00338	17818	19973

Results are from ACF specification that allows for differing capital and labor coefficients by SOE status.

In terms of sign, the results seemingly confirm the hypothesis that SOEs adopt more capital intensive technologies. Almost every 2-digit industry has a higher labor coefficient for SOEs, as evidenced by the positive values for γ_l. Additionally, the industries with more negative values for γ_k tend to have more positive values for γ_l. More formally, I added the two γ values together for each industry, and the mean (weighting each industry equally) change in returns to scale was about -.02. While not exactly 0, this suggests that the alternative specification is not uncovering dramatically different returns to scale for SOEs vs. private firms. At any rate, it is smaller in magnitude than the vast majority of changes in either coefficient. Between the two γ values generally taking the opposite sign and returns to scale staying about the same across specifications (for both types of firms), it is safe to say that the SOE specification indicates that private firms tend to use more capital intensive technologies than state-owned firms.

A Note on Standard Errors and OLS

The full set of standard errors for the baseline and SOE specifications is not included in this iteration of the paper. Ultimately, every industry's estimates in both specifications will be bootstrapped. In some preliminary testing on four sample industries, the results were fairly significant based on 150 iterations. In all four industries, whose values for γ_k were near the overall mean observation, a positive value for γ_k was never realized, while γ_l was negative less than 10% of the time in all four industries, and less than 5% of the time in three of the four. Thus, while it is impossible to assert the significance of the results for every industry individually at this time, it appears that the general result–SOEs have lower capital coefficients and higher labor coefficients–will be statistically significant.

For a quick basis of comparison, the interaction terms were significant in all but six industries when this same model was run using OLS. Private capital coefficients tended to be smaller in the OLS specification by about .05 to .10, which is an expected bias that the ACF framework is meant to correct. Similarly, labor coefficients decreased by a similar magnitude between OLS and ACF. To the extent that OLS is biased and the ACF corrections are important, these differences appear to justify the choice of a more complex estimator to recover the production function coefficients.

These results have been generated by a just identified system–that is, only contemporaneous labor and capital (and SOE) values were used as instruments. Under the ACF assumptions, lagged values of materials, capital, and labor are also permissible as instruments to achieve overidentification. I have experimented with these additional instruments in alternative specifications, and while some of the differential terms get smaller, the general results do not appear to change.

3.5.4 Baseline Productivity Results

With our testable assumptions satisfied and our production function estimates looking reasonable, we can turn to analyzing our implied productivity residuals. For a baseline to compare against, I ran the Pakes and Olley (1996) decomposition (OP) across all industries on the vanilla Ackerberg et al. (2005) specification. The OP decomposition works in the following way:

1. Let $\Phi_t = \sum_{i=1}^{N} s_{it}\omega_{it}$, a share-weighted [12] measure of aggregate productivity.
2. Decompose it into two terms: $\Phi_t = \bar{\Phi}_t + \sum_{i=1}^{N}(s_{it} - \bar{s}_{it})(\omega_{it} - \bar{\omega}_{it})$

The first term, $\bar{\Phi}_t$, is the unweighted mean of productivity in the sample. It is meant

[12] In this case, shares are calculated with value-added.

to capture "organic" TFP growth across all firms. If all firms grew by the same amount in TFP, $\bar{\Phi}_t$ would correspondingly grow the same amount, while this is not necessarily true for the second term.

The second term is meant to capture the extent to which allocation to more efficient firms contributes to overall productivity. That is, if more productive firms have higher market shares, then aggregate productivity will be higher, and this term isolates that effect from standard firm-level TFP growth.

Table 3.6: Olley-Pakes Decomposition - Aggregate Baseline

Year	Aggregate	Unweighted Mean	Allocation	N
1998	1.386	1.910	-0.524	103840
1999	1.478	1.948	-0.470	105344
2000	1.552	1.989	-0.438	107949
2001	1.590	2.059	-0.469	118413
2002	1.672	2.129	-0.457	130809
2003	1.792	2.194	-0.403	146849
2004	2.071	2.219	-0.148	225762
2005	2.177	2.235	-0.0576	224502
2006	2.241	2.270	-0.0295	250124
2007	2.340	2.346	-0.00627	230487

Results are from standard ACF specification that does not allow for differing production functions.

It is plain to see from Table 3.6 that aggregate productivity is growing for much of the sample. The TFP measure more than doubles over the sample period, as does the number of firms [13].

While the "unweighted" component is larger in magnitude, the aggregate TFP growth in China over the sample period is roughly due in equal parts to unweighted growth and allocation growth. That is, firms in the market generally increase their TFP over time, and more efficient firms are seeing increasingly high market shares.

[13] Any discrepancies from 3.1 are due to either missing values of key variables, or because some smaller industries did not have enough observations to be estimated.

In performing the OP decomposition on the aggregate economy, we have a basis of comparison if we do the same thing on SOEs only. Again, while it is difficult to compare TFP residuals across industries and subsamples, their residuals that went into calculating the previous table will still be the same, and we are now just excluding private firms and re-weighting the shares accordingly.

Table 3.7: Olley-Pakes Decomposition - SOE Baseline

Year	Aggregate	Unweighted Mean	Allocation	N
1998	0.774	1.223	-0.449	33351
1999	0.826	1.250	-0.425	31782
2000	0.831	1.273	-0.442	29745
2001	0.784	1.304	-0.520	27678
2002	0.807	1.351	-0.544	26981
2003	0.874	1.404	-0.530	26397
2004	1.163	1.520	-0.357	33166
2005	1.251	1.552	-0.301	31005
2006	1.224	1.595	-0.371	30965
2007	1.291	1.743	-0.452	26482

Results are from ACF specification that restricts production functions to be the same. Sample is SOEs only.

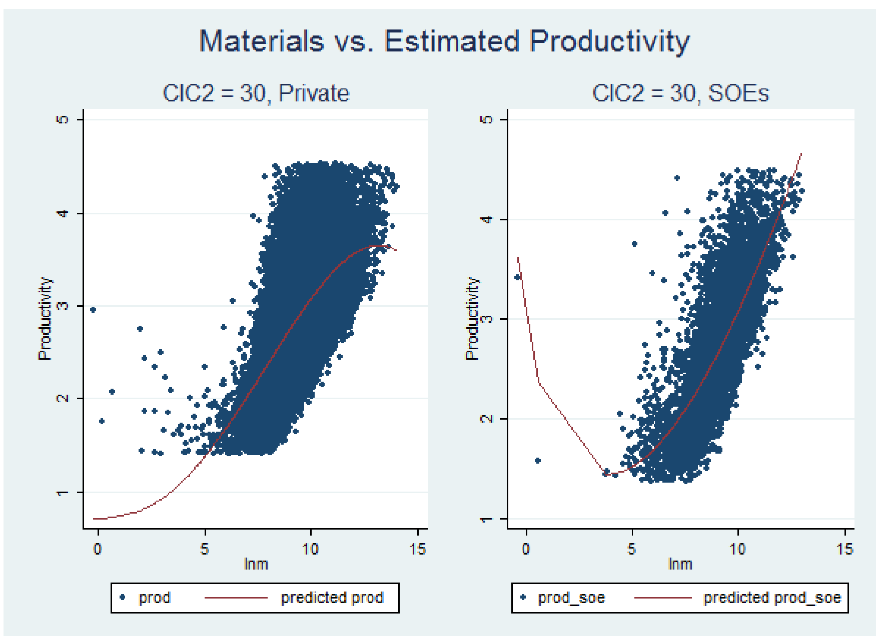
Table 3.7 shows that the SOE-only sample skews even more toward unweighted growth. In fact, the unweighted mean of TFP of all firms by 2007 is twice as high as the aggregate TFP measure in 1998. While there is some allocation growth through the sample, it is not nearly as big as a contributor as with private firms. This makes some amount of sense–to the extent that SOEs may face less competitive markets, a state-owned firm may not have to be as efficient to survive and grow. In terms of percentage growth, SOEs are doing better than the aggregate economy. However, without the major drag generated by the allocation term, they would be even more efficient.

This table also helps to illustrate that SOEs are generally less productive than private firms despite this faster growth. Both weighted and unweighted TFP is smaller in every year of the sample. However, the unweighted SOE-mean TFP estimate is quickly catching up to the overall estimate, growing by over 40% in the sample period as compared to the overall rate of about 23%. Despite this, there is still a sizable gap between the two types of firms as of 2007.

3.5.5 Initial SOE Specification Results

Before comparing the results from the SOE specification to the baseline specification, we need to check our monotonicity assumption yet again. This time, since we have allowed for separate technologies and policy functions for both types of firm, we need to check it in each case:

Figure 3.2



For this sample industry, it appears that the assumption holds at least as well as it did in the baseline specification for both types of firm, save for some outliers in the lower tails [14]. We can now run the TFP residuals generated from the SOE specification through the Olley-Pakes decomposition. Since we have essentially doubled the number of technologies present in the economy, we cannot directly compare even the aggregate results–they represent a fundamentally different set of production functions. Even only comparing SOEs or only comparing private firms, the numbers are in some sense on a different scale.

One advantage of the decomposition is it provides concise summaries of different components of aggregate productivity. While the terms are not always directly comparable due to differences in production functions, it is possible to look at the relative magnitude of different terms across specifications. To start, take the OP decomposition applied only to SOEs given their new technologies:

Table 3.8: Olley-Pakes Decomposition - SOEs Only (SOE Specification)

Year	Aggregate	Unweighted Mean	Allocation	N
1998	0.832	1.234	-0.402	33351
1999	0.883	1.265	-0.383	31782
2000	0.898	1.288	-0.391	29745
2001	0.846	1.318	-0.472	27678
2002	0.873	1.365	-0.492	26981
2003	0.951	1.421	-0.469	26397
2004	1.236	1.513	-0.276	33166
2005	1.339	1.551	-0.213	31005
2006	1.310	1.595	-0.284	30965
2007	1.389	1.749	-0.360	26482

Results are from ACF specification that allows production functions to differ. Sample is SOEs only.

At first glance, it looks like allocation has gone up, while the unweighted means have barely changed. Given that this decomposition represents a different set of technologies, we need to cast this analysis in more directly comparable numbers. Below is a table that takes the difference between aggregate productivity and the allocation term, and

[14]I have checked with several other industries and the assumption appears to hold in general.

normalizes it relative to aggregate productivity that year:

Table 3.9: Relative Deviations of Allocation Term - SOEs

Year	Baseline	SOE
1998	-1.580	-1.483
1999	-1.515	-1.434
2000	-1.531	-1.436
2001	-1.663	-1.558
2002	-1.673	-1.563
2003	-1.607	-1.494
2004	-1.307	-1.223
2005	-1.240	-1.159
2006	-1.303	-1.217
2007	-1.350	-1.259

We can see that, when appropriately normalized by the implied levels of aggregate SOE productivity, the switch in specifications generates a reduction in the drain caused by "misallocation" by 8% to 11% each year. While this is not an overwhelmingly large change induced by the SOE specification, a story is beginning to emerge: if one accounts for the differing forms of technology adaptation induced by state ownership in China, then the negative role of misallocation among SOEs is mitigated.

We can also check this result for private firms only:

Table 3.10: Relative Deviations of Allocation Term - Private

Year	Baseline	SOE
1998	-1.007	-1.025
1999	-0.986	-1.002
2000	-0.956	-0.971
2001	-0.950	-0.965
2002	-0.943	-0.960
2003	-0.921	-0.938
2004	-0.917	-0.938
2005	-0.898	-0.915
2006	-0.887	-0.904
2007	-0.885	-0.902

The result, in light of the findings for SOEs, makes sense: one group's level of misallocation was overstated, while the other's was understated. Given that there are many more private firms, and they thus had a higher level of influence on the baseline estimates, it makes sense that the effect is more muted for them. With the much larger number of private firms in the economy (especially in later years), a 2% difference could have meaningful welfare implications.

3.5.6 Dispersion

There are a few possible sources of this decline in misallocation. It could be that the relative rankings of shares and productivities change across specifications. A firm's share will be the same in either the baseline or SOE specification, so the only possible mechanism for changing rank is on the productivity end. The correlation between the two sets of TFP residuals is over .98 [15], which indicates that there are not a massive number of firms shifting around their ranks in the productivity distribution.

This suggests that the shape and/or dispersion level of the productivity distribution is somehow changing across the two specifications. It is difficult to compare the two distributions directly because they reflect the residuals of different production functions. For example, since the SOE specification has more explanatory variables in its function, we would expect the variance of the implied residuals to be smaller [16]. So, it is necessary to find a metric that takes into account the fact that these residuals are coming from distributions with different scales. One option is to take the ratio of different percentiles– this essentially normalizes the residuals with respect to scale and gives us a comparable metric.

[15]This result is very robust to different samples and the exclusion of outliers.

[16]While ω_{it} is a different object than a traditional regression residual, this finding seems to apply to both ω_{it} and ε_{it}.

It makes the most sense to compare dispersion levels at the industry/ownership level across specifications. One candidate industry, plastic products manufacturing, is outlined below [17]:

Table 3.11: Dispersions Across Specifications - SOEs Only

Year	CIC2	75-25 Ratio	75-25 Ratio, SOE	90-10 Ratio	90-10 Ratio, SOE
1998	30	1.320	1.477	1.737	2.014
1999	30	1.343	1.465	1.718	2.025
2000	30	1.370	1.497	1.746	2.059
2001	30	1.316	1.413	1.708	1.977
2002	30	1.339	1.433	1.727	1.976
2003	30	1.337	1.410	1.613	1.844
2004	30	1.282	1.378	1.598	1.791
2005	30	1.279	1.370	1.615	1.797
2006	30	1.245	1.347	1.602	1.831
2007	30	1.264	1.365	1.594	1.818

Both the 75-25 and 90-10 measure of dispersion increase in the SOE specification. The 90-10 dispersion ratio is particularly strong: in 1999, the 90th percentile is on average 172% of the 10th percentile in the baseline specification, and this figure increases to 203% in the SOE specification. The productivity values in the right tail have gotten substantially larger than those in the left, relatively speaking. This poses an additional question: is this effect due to the right tail getting larger, the left tail getting smaller, or both? To examine this question, we can look at a couple of other percentile ratios:

[17] This industry was chosen because its results are generally representative. I have run this analysis on every industry, and with the exception of about 4 of them the results take on the same signs, and the magnitudes are at least as large.

Table 3.12: Dispersions Across Specifications - SOEs Only

Year	CIC2	90-50 Ratio	90-50 Ratio, SOE	50-10 Ratio	50-10 Ratio, SOE
1998	30	1.318	1.418	1.318	1.421
1999	30	1.311	1.424	1.311	1.422
2000	30	1.324	1.439	1.319	1.431
2001	30	1.272	1.374	1.342	1.439
2002	30	1.283	1.365	1.346	1.448
2003	30	1.255	1.314	1.285	1.404
2004	30	1.266	1.338	1.263	1.339
2005	30	1.245	1.323	1.298	1.359
2006	30	1.248	1.328	1.284	1.378
2007	30	1.257	1.355	1.268	1.341

At first glance, this table indicates that any possible increase in dispersion is due roughly in equal part to the expansion of the right and left tails. This is not quite true, since the expansion in the 90-50 ratio already incorporates the prior expansion in the 50-10 ratio. So, taking measures that move across the whole distribution like our misallocation term, the right tail will be much larger in magnitude than it was before if we increase the 90-50 and 50-10 ratios the same amount, while the median will only be modestly larger. Thus, this table provides mild support for the hypothesis that the results thus far are disproportionately driven by the right tail expanding. Given that the unweighted mean remains more or less unchanged across specifications, it appears that it may be due to productivity increases in just a few large state-owned firms.

For a point of comparison, we can also look at the dispersion changes in privately owned firms:

The corresponding relative lack of change in dispersion in private firms tracks with the earlier result on misallocation.

Table 3.13: Dispersions Across Specifications - Private Only

Year	CIC2	75-25 Ratio	75-25 Ratio, SOE	90-10 Ratio	90-10 Ratio, SOE
1998	30	1.261	1.263	1.612	1.610
1999	30	1.281	1.274	1.619	1.605
2000	30	1.268	1.264	1.588	1.579
2001	30	1.263	1.260	1.583	1.568
2002	30	1.264	1.264	1.587	1.580
2003	30	1.258	1.254	1.580	1.570
2004	30	1.251	1.244	1.548	1.539
2005	30	1.254	1.246	1.562	1.542
2006	30	1.263	1.256	1.571	1.555
2007	30	1.266	1.259	1.580	1.567

3.5.7 Entry and Exit

Up to this point, the analysis has not exploited the panel dimension of the data except in the ACF estimation routine. But, as can be seen from the summary statistics, there is massive entry of private firms into the economy as well as a corresponding exit of SOEs over the sample period. Entry and exit margins can be key determinants of aggregate productivity for both of our subsamples of concern. The OP decomposition, while useful, does not get at the relative contributions of entering and exiting firms. So, it is necessary to use a more complex decomposition to separate all of the different groups who can influence aggregate productivity.

Melitz and Polanec (2013) Decomposition

The Melitz and Polanec (2013) (known as MP) decomposition seeks to quantify the contributions of a change in productivity over two periods (ie, year 1 and year 2). Firms get classified into 3 groups:

1. Firms who are present in periods 1 and 2 are "survivors" (group S).

2. Firms who are present in period 1 and not 2 are "exiters" (group X).

3. Firms who are present in period 2 and not 1 are "entrants" (group E).

In the context of the two periods, these three groups are mutually exclusive and exhaustive. The MP decomposition features Φ as its measure of aggregate productivity, which is the same from the OP decomposition.

Classifying firms into three groups and getting the relative group shares and Φ's results in the following identities (X denotes exit, E denotes entrant):

$$\Phi_1 = s_{S1}\Phi_{S1} + s_{X1}\Phi_{X1} \tag{3.10}$$

$$\Phi_2 = s_{S2}\Phi_{S2} + s_{E2}\Phi_{E2} \tag{3.11}$$

Note that the group-specific Φ values use inside shares, or else the identity would not hold. Through some basic algebra and substitution comes the basic form of the decomposition:

$$\Delta\Phi = (\Phi_{S2} - \Phi_{S1}) + s_{E2}(\Phi_{E2} - \Phi_{S2}) + s_{X1}(\Phi_{S1} - \Phi_{X1}) \tag{3.12}$$

In this decomposition, the identity is applied to changes in aggregate productivity over time. The first parenthetical term is the net contribution of the change in productivity of surviving firms, while the next two capture entrants and exiters. The "survivor" term can be further decomposed into an unweighted and allocation term, like in OP, if desired.

To begin, I ran the MP decomposition on the full set of SOEs, using the baseline specification:

Table 3.14: MP Decomposition - SOE Only

Year	$\Delta\Phi$	Survivors	Entrants	Exiters
1999	0.0511	0.0379	0.0810	-0.0678
2000	0.00578	0.0619	-0.0218	-0.0344
2001	-0.0474	-0.0630	0.0931	-0.0776
2002	0.0232	0.0382	0.0606	-0.0755
2003	0.0667	0.0946	0.0712	-0.0991
2004	0.289	0.219	0.200	-0.130
2005	0.0888	0.0886	0.120	-0.120
2006	-0.0275	-0.00197	0.0583	-0.0838
2007	0.0670	0.0914	0.0936	-0.118

Results are from ACF specification that restricts production functions to be the same. Sample is SOEs only.

The $\Delta\Phi$ terms match up with the changes in the "Aggregate" term from year to year in the comparable OP decomposition table. Just looking at SOEs, it appears that all three groups have roughly comparable contributions to aggregate productivity changes, in terms of magnitude. However, this table is misleading. In taking the MP decomposition and applying it to SOEs only, there is an implicit assumption in only using this categorization that being an SOE is a permanent state. An SOE who turns private gets listed as an exiter, when really they remain in the economy.

To address this issue, a modification to the decomposition is necessary. Consider, instead of all SOE observations in a two year period, all firms who were an SOE in either or both periods. Next, define the following five groups:

1. Firms who are SOEs in periods 1 and 2 are "survivors" (group S).
2. Firms who are SOEs in period 1 and shut down in period 2 are "exiters" (group X).
3. Firms who are SOEs in period 2 and closed in period 1 are "entrants" (group E).

4. Firms who are SOEs in period 1 and operate privately in period 2 are "privatized" (group P).
5. Firms who are private in period 1 and SOEs in period 2 are "nationalized" (group N).

This is an exhaustive and mutually exclusive categorization of all firms that operated as SOEs at any point over two periods. Via similar identities and algebra as before, the new decomposition takes the following form:

$$\Delta\Phi^{SOE} = (\Phi_{S2}^{SOE} - \Phi_{S1}^{SOE}) + s_{E2}(\Phi_{E2}^{SOE} - \Phi_{S2}^{SOE}) + s_{X1}(\Phi_{S1}^{SOE} - \Phi_{X1}^{SOE}) + s_{P1}(\Phi_{S1}^{SOE} - \Phi_{P1}^{SOE}) + s_{N2}(\Phi_{N2}^{SOE} - \Phi_{S2}^{SOE}) \tag{3.13}$$

When applied to private firms, the roles of nationalization and privatization need to be switched. The modified MP decomposition leads to the following results on all SOEs in the baseline specification:

Table 3.15: MP Decomp. with Privatization - SOE Only

Year	$\Delta\Phi$	Survivors	Entrants	Exiters	Privatized	Nationalized
1999	0.0511	0.0379	0.0597	-0.0498	-0.0180	0.0214
2000	0.00578	0.0619	-0.0406	-0.00865	-0.0257	0.0189
2001	-0.0474	-0.0630	0.0551	-0.0513	-0.0262	0.0380
2002	0.0232	0.0382	0.0290	-0.0352	-0.0403	0.0316
2003	0.0667	0.0946	0.0241	-0.0487	-0.0504	0.0471
2004	0.289	0.219	0.117	-0.0414	-0.0889	0.0825
2005	0.0888	0.0886	0.0308	-0.0460	-0.0742	0.0896
2006	-0.0275	-0.00197	-0.0124	0.0198	-0.104	0.0706
2007	0.0670	0.0914	0.00754	-0.0156	-0.102	0.0860

Results are from ACF specification that restricts production functions to be the same. Sample is SOEs only.

As established in the OP decomposition, SOE aggregate productivity is not particularly volatile from year to year. This table shows that this is not necessarily due to a lack of volatility at important margins–entry and exit generally take opposing signs, as do privatization and nationalization. Thus, the stable numbers for $\Delta\Phi$ are actually covering up large amounts of "churn" from period to period.

The signs of the terms on survivors and entrants are mostly as expected. Given that there is established TFP growth among SOEs already, it is reasonable that this extends to incumbent firms rather than just along entry/exit esque margins. Similarly, the entry margin is usually positive. This means that when a new state-owned corporation is started, it is generally more productive than the average. While this is not traditionally true for private firms in China, since SOEs are being started by a central planning entity with immense resources, it is fair to expect these firms to do better upon entry than their private counterparts. The same is true for the nationalization term–if, despite the mass exit of SOEs in the economy, the government decides to nationalize a firm, it could either be regionally prominent or a notable success story. It is possible that the government would look to "resuscitate" struggling private firms (which may give us a negative sign on the nationalization term), but since SOEs are generally at a TFP disadvantage anyway, this may not prove a successful strategy.

More surprising, however, are the exiter and privatization results. A generally negative exiter sign suggests that, on average, SOEs that are shut down are generally more productive than the average incumbent firm. While TFP is not the sole metric of a firm's success, it stands to reason that more efficient firms would be kept open. It is easier to rationalize the sign on the privatization term–while the government is actively making average SOE productivity go down, it may be beneficial from a central planning perspective for these particular firms to be put into the private sector and face increased competition. This is slightly at odds with the story behind the nationalization coefficient: if the government wants the private sector to be filled with efficient firms, they would not be removing them via nationalization. The explanation for this likely falls along the same lines as Hsieh and Song (2015): there are generally very few firms nationalized, and they are on average very large, while the privatized firms likely represent a larger cluster of smaller firms.

A couple of caveats are necessary with this table. As mentioned earlier, the sample cleaning process has disproportionately affected exiting SOEs. So, it stands to reason that this decomposition may be particularly sensitive to the choices made in sample selection. Robustness checks involving alternative sampling schemes will be included in a future iteration of this paper.

There are now five possible terms for the SOE specification to change, rather than just two:

Table 3.16: MP Decomp. with Privatization - SOE Only (SOE Specification)

Year	$\Delta\Phi$	Survivors	Entrants	Exiters	Privatized	Nationalized
1999	0.0508	0.0390	0.0632	-0.0553	-0.0186	0.0225
2000	0.0149	0.0674	-0.0358	-0.0104	-0.0266	0.0203
2001	-0.0516	-0.0662	0.0540	-0.0520	-0.0270	0.0396
2002	0.0274	0.0431	0.0297	-0.0366	-0.0414	0.0326
2003	0.0779	0.103	0.0272	-0.0502	-0.0515	0.0494
2004	0.285	0.212	0.127	-0.0476	-0.0909	0.0854
2005	0.102	0.0976	0.0354	-0.0479	-0.0754	0.0924
2006	-0.0282	-0.00113	-0.0174	0.0261	-0.106	0.0701
2007	0.0785	0.0992	0.00733	-0.0112	-0.104	0.0872

Results are from ACF specification that allows production functions to differ by state ownership. Sample is SOEs only.

While this table is not re-normalized for comparison with the baseline specification, it is already plain to see that almost no term has changed meaningfully in magnitude or sign. This is in some sense a null result–incorporating technological differences between SOEs and private firms does not change any of the basic story that this particular decomposition tells. But, it does serve as a robustness check on the fairly surprising results in the baseline specification. Neither the unexpected signs on privatization nor exiters are changed [18].

Given the modest results for SOEs, and the fact that results for private firms have been more modest thus far, the MP decomposition tables for private firms have not been included

3.6 Extensions

3.6.1 Extension: Balat and Sasaki (2014) meets Levinsohn and Petrin (2003)

The final set of results relies on a slightly more complex model, from Balat and Sasaki (2014), with a production function that differs even more at the firm level:

$$y_{it} = \beta_l^{it} l_{it} + \beta_k k_{it} + \omega_{it} + \varepsilon_{it} \tag{3.14}$$

While only the labor coefficient has been allowed to differ at the firm level, this makes identification and estimation significantly more complicated. First, it is necessary to modify the set of assumptions from our original model. The new assumptions are as follows:

1. There are now two proxy variables that play an equivalent role to intermediate inputs in the prior models. We denote this as $M_{it} \in \mathbf{R}^2$.

[18]Given the modest results for SOEs, and the fact that results for private firms have been more modest thus far, the MP decomposition tables for private firms have not been included in the paper.

2. There is now a direct invertibility assumption on the firm-level heterogeneous terms and the proxy variables: $(\beta_l^{it}, \omega_{it}) = g_t^{-1}(k_{it}, m_{it})$.

3. The labor market is subject to frictions, such that there are idiosyncratic shocks (denoted ε_{it}^{l}) that influence the choice of labor separately from the choices of the proxy variables and capital.

4. ε_{it} and ω_{it} behave the same as in prior models otherwise.

Under the new invertibility assumption, one can write the model the following way:

$$y_{it} = \beta_l^{it}(k_{it}, M_{it})l_{it} + \beta_k k_{it} + \omega_{it}(k_{it}, M_{it}) + \varepsilon_{it} \tag{3.15}$$

Or:

$$y_{it} = \beta_l^{it}(k_{it}, M_{it})l_{it} + \Phi_{it}(k_{it}, M_{it}) + \varepsilon_{it} \tag{3.16}$$

Stage 1: Labor Productivity

Since ε_{it} is fully exogenous and mean 0 by assumption, we also know the following:

$$E[y_{it}|k_{it}, l_{it}, M_{it}] = \beta_l^{it}(k_{it}, M_{it})l_{it} + \Phi_{it}(k_{it}, M_{it}) \tag{3.17}$$

The heterogeneous labor coefficient is identified from this equation, according to Balat and Sasaki (2014). If we are able to estimate the conditional mean function non-parametrically from this equation, then its derivative with respect to labor will return the coefficient. This is very similar to the first stage of Pakes and Olley (1996) or Levinsohn and Petrin (2003), but the estimation is now more complicated than just using OLS to recover a labor coefficient that does not differ across firms or time [19].

To estimate this coefficient, I borrow the "local derivative" approach from Balat and Sasaki (2014). First, we assume the regression function is to be estimated via a kernel regression:

$$\hat{E}[y_{it}|k_{it}, l_{it}, M_{it}] = \frac{\sum_{j' \neq j} y_{j't} w_{jj't}}{\sum_{j' \neq j} w_{jj't}} \tag{3.18}$$

Where for firms j and j', $w_{jj't}$ refers to the following object:

$$w_{jj't} = \kappa(\frac{k_{j't} - k_{jt}}{b_k})\kappa(\frac{m_{1j't} - m_{1jt}}{b_{m1}})\kappa(\frac{m_{2j't} - m_{2jt}}{b_{m2}})\kappa(\frac{l_{j't} - l_{jt}}{b_l}) \tag{3.19}$$

κ is, in this case, the standard Gaussian kernel function, and b_k, b_{m1}, b_{m2}, and b_l are bandwidth parameters [20]. Rather than estimating this object directly, it is necessary to apply the quotient rule with respect to l_{it}:

$$\hat{\beta}_l^{it}(k_{it}, M_{it}) = \frac{\partial \hat{E}[y_{it}|k_{it}, l_{it}, M_{it}]}{\partial l_{it}} = \frac{\hat{bf}_{it}\hat{bg}_{lit} - \hat{bf}_{lit}\hat{bg}_{it}}{\hat{bf}_{it}^2} \tag{3.20}$$

[19] These first stages suffer from a collinearity problem, as pointed out by Ackerberg et al. (2005), where firms make their labor decisions based on the same information as the flexible input decisions. The additional assumption on ε_{it}^l is a way to circumvent this critique, where labor decisions are subject to idiosyncratic variation as compared to the intermediate decisions

[20] For now, I have chosen preliminary bandwidths in a fairly simple manner. I have applied the multivariate version of Silverman's rule of thumb, and then inflated the bandwidths by 3 so as to account for the fact that when estimating a density derivative, the estimation variance is much more sensitive to increases in bandwidth. While this is not an optimal solution, I plan to implement least-squares cross-validation eventually to get better estimates.

Where:

$$\hat{bf}_{it} = \sum_{j' \neq j} w_{jj't} \tag{3.21}$$

$$\hat{bg}_{it} = \sum_{j' \neq j} y_{j't} w_{jj't} \tag{3.22}$$

$$\hat{bf}_{lit} = \sum_{j' \neq j} \frac{\partial w_{jj't}}{\partial l_{it}} \tag{3.23}$$

$$\hat{bg}_{lit} = \sum_{j' \neq j} y_{j't} \frac{\partial w_{jj't}}{\partial l_{it}} \tag{3.24}$$

Stage 2: Capital Coefficient

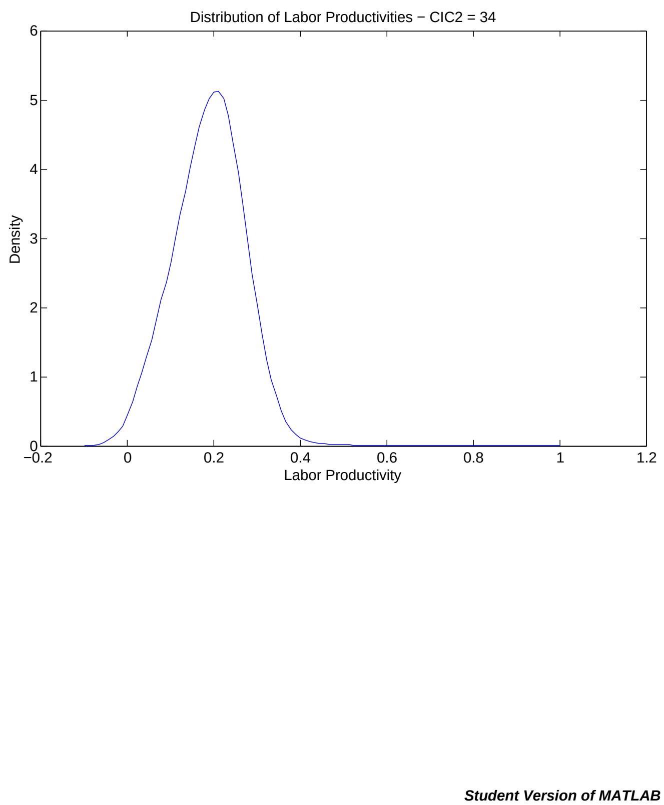
With the labor productivity estimates, define a new variable, $\tilde{y}_{it} = y_{it} - \hat{\beta}_l^{it} * l_{it}$. With $\tilde{y}_{it}$, we can create a second stage that is analogous to Pakes and Olley (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2005). Rewrite the model as:

$$\tilde{y}_{it} = \Phi_t(k_{it}, M_{it}) + \varepsilon_{it} \tag{3.25}$$

Given our assumptions on ω_{it}. With non-parametric estimates of Φ_t, our second stage can proceed exactly like the one in Ackerberg et al. (2005), only with $\tilde{y}_{it}$ as the dependent variable, and only searching over candidate capital coefficients. With both labor and capital coefficients, we can recover ω_{it} as in our earlier specifications using $\hat{\Phi}_t$.

3.6.2 Preliminary Results

Note: there are still some errors or issues with my implementation of this estimator. For now, I have just included some preliminary graphs and summary statistics from what I have gotten. "Financial expenses" was chosen as a second proxy variable (to get the dimension of M_{it} up to 2). The reasoning behind this is similar to that in the choice of intermediate materials: increases in productivity, all else equal, will generally induce increases in production. This increased scale will usually lead to a higher use of intermediate materials as well as financial expenses even though neither term enters the value-added production function directly. Thus, the monotonicity assumption should also apply to financial expenses despite their not being a traditional intermediate input like fuel or energy. The following graph is a kernel density of the estimated labor productivities in CIC2 34 from 1998 to 2007:



Any estimates below -.2 and above .95 have been removed for this graph, though this cleaning resulted in the removal of under 20 observations out of over 30,000 used in the final estimation. The mean estimate of β_l^{it} is roughly .145–this is far lower than the baseline estimate, or either of the labor coefficients implied by the SOE specification (or OLS, for that matter). The implied capital coefficient is about .13, which is similarly problematic. As such, this method suffers from either an identification issue or a coding error at this point, and I will refrain from interpreting the results in this iteration of the paper. With such small coefficients (and an implied mean return to scale of under .4), the TFP residuals will be far larger than any of the specifications so far.

3.7 Conclusion and Future Research

This paper has documented, via structural estimates done in the (Ackerberg et al., 2005), that state-owned firms in China operate significantly different technologies than their private counterparts in the same industry. Specifically, they are almost uniformly less capital intensive and more labor intensive, likely owing to the fact that capital is more freely and easily available to these firms. The ensuing lack of competition for these firms in capital markets may have lead them to "overconsume" capital, thus depressing marginal products of capital below where they would otherwise be.

When accounting for these technological differences in various productivity analysis, it has been found that they have a meaningful impact on two dimensions: the drain of misallocation on measure of aggregate TFP is reduced, and within-SOE productivity dispersion is substantially increased. The margins of SOE entry, exit, privatization and nationalization are found to be important in explaining aggregate SOE TFP patterns. But, these margins are already so large and volatile that incorporating technological differences does not change the overall story.

Preliminary examinations of fully heterogeneous firms, where every firm in the economy is allowed to operate a technology with different Cobb-Douglas coefficients, suggests that there is still-uncovered meaningful dispersion across firms at this level. Incorporating such estimates into the analysis above is a focus of future work.

There are some notable shortcomings to this analysis. This paper suffers from a flaw that plagues much of the productivity estimation literature: prices and quantities are not separately observed for key variables. This means that demand-side factors, or in the case of China, explicit pricing policies often get absorbed into the TFP residual. Given that the same prices are unobserved in both specifications, their effect would at least be constant when comparing the results of this paper.

Luckily, monthly quantity data for output has recently become available for 2000-2006. In the vein of Foster et al. (2008), having this data will open several new avenues of analysis when demand influences can be isolated separately from pure technical efficiency. Similarly, revenue TFP and quantity TFP are important objects in (Hsieh and Klenow, 2009), though they are backed out in the context of a macro model rather than observed directly in that paper.

While this paper is able to determine descriptive patterns of different technologies used by different firms, these ideas could be better applied to a full-fledged structural model of technology adoption at the firm level. This could lead to more refined counterfactual analysis of the patterns of state ownership of firms. That is, if one could model the firm's decision to choose a discrete technology, then a structural model that evaluates the gains and losses of the favors and advantages given to state-owned firms could be created.

APPENDIX A

CHAPTER 1 OF APPENDIX

A.0.1 Physical Sample Construction

The largest possible sample from my overlapping datasets comes from the physical census of powerplants. Pre-cleaning, there are 21,100 observations in total, with 4,136 unique power plants. The median span a plant is observed is just 4 years, which is likely a combination of the high entry and exit seen in this dataset, as well as because of extensive measurement error in key plant identifier variables. Despite this, there are a fair number of plants with uninterrupted observation spells (post entry at least). 113 Are present for the entire sample going back to 1995, while over 860 enter at a later year and are present for the rest of the sample. Plants who are present from the start of the sample and appear to have "genuine" exits add another 360 or so observations. All observations are missing 1996, 1999, and 2001. 1998 is a more relevant starting point as this is where any and all financial data is available, in which case 208 plants are observed for the entire sample.

The first modification I make is to drop all "captive" power plants, which do not supply power to the electric grid, but instead transmit directly to one client (usually a large industrial plant that also owns the power plant). This lowers the total number of observations to 18166, and the number of unique power plants to 3,080. Now only 81 power plants are present for the entire sample, while more like 755 plants have long and uninterrupted spells (longer than 3, present for the entire sample period post entry). "Genuine exit" plants now constitute another 235 observations.

Finally, I drop all power plants whose maximum observed capacity is less than 50

MW. These plants are extremely small compared to most of the players in the market, and are generally understood to operate by a different set of regulations because of this. In sections of my analysis that require using total output, I use aggregate figures that include their production. This halves the sample to 7,993, though the amount of aggregate capacity that is retained is well above 70%. There are now 50 plants present for the entire sample, 1,471 unique plants, and roughly 522 "uninterrupted" plants in total with 108 "genuine exit" plants.

A.0.2 Revenue Sample(s) Construction

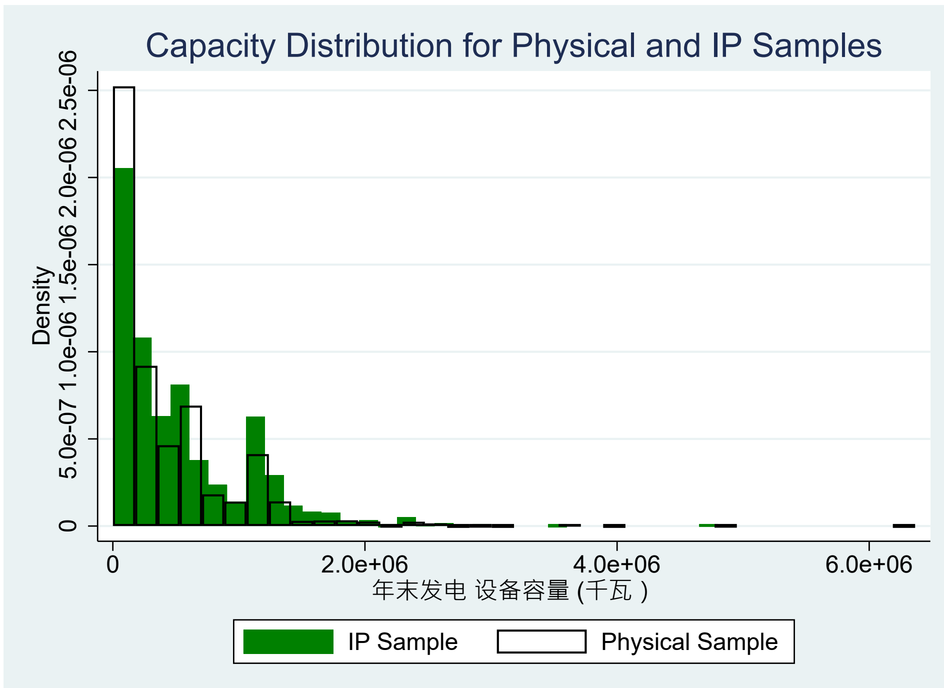
I also drop plants that are anomalous in terms of heat rate, input price, and output price (equivalent to roughly winsorizing at about 3 or 4% at the moment). This leaves me with a total of 1,960 observations that have all possible relevant variables.

The key observations for this paper will be those that merge successfully with the NBS financial data. plants get featured in this data when they are above a certain revenue threshold (5 million RMB from 1998 to 2011, 20 million RMB after 2011), though due to possible measurement errors I do not assume that a plant's absence from the NBS sample necessarily means it is below this cutoff.

Output price information is available for 2,371 observations from 1998 to 2009 (with 1999 and 2001 missing physical data). Input prices, which are a key component of marginal cost calculations, are only available through 2007 and with 2004 added, resulting in 2,069 observations. Given that my analysis for now makes very broad assumptions about output prices, I work with the limited input price sample for my first stage analysis (and an even further limited sample for the dynamic analysis).

A.0.3 Sample Comparisons

Since the census data tends to cover larger plants, it skews slightly more toward the right tail than the physical sample. However, the samples do not have radically different capacity distributions:



A.0.4 Import/Export Data Imputation

For exports, I observe 521 of the possible 600 province-year combinations. Missing observations may be due to measurement error or due to genuine "zeros" where a province did not export any electricity. In cases where this distinction is ambiguous, the total amount exported would likely be small enough so as not to substantially change my analysis.

For imports, there are 493 observations. So, it is necessary to impute trade balance data for fully missing observations, as well as those where one of the two trade variables is missing.

To impute these missing trade balance values, I impute missing import and export values according to the following process:

1. If there are observations before and after a missing value, use linear interpolation.
2. If a missing observation occurs before (or after) all import and export information, and both are available, get the ratio of imports to exports in the earliest (latest) year where that is observed.
3. If one of imports or exports is observed and the other is missing, apply this ratio to impute the missing value.
4. If both imports and exports are missing for observations before or after all other import and export information, assume a "real" 0.

A.0.5 Dynamic State Space Approximation

Even exploiting the computational simplicity of Rust (1987), a 5-dimensional model is likely unstable, and how plants treat the aggregate states needs to be specified.

The full revenue function for each plant depends on μ, cap, mc, Q, p, and every other plant's μ. I assume that plants forecast the evolution of μ and mc under uncertainty, but have perfect foresight over both market demand and other macro shocks, as represented by market and year dummies. Thus there is now a unique EV_{mt} function to be solved for each province-year combination.

Based on the components of the static portion of the model, plants face the following revenue functions (including year and individual fixed effects):

$$\frac{q_{it}}{cap_{it}} = exp(\beta_0 + \beta_1 cc_{it} + \beta_i + \beta_t + \mu_{it} - ln(s_0(Q_{mt}, \mu_{j\neq i}))) \tag{A.1}$$

$$q_{it} = s_{it} cap_{it} \tag{A.2}$$

and

$$p_{it} = \alpha_0 + \alpha_1 cc_{it} + \alpha_2 ln(cap_{it}) + \eta_{it} \tag{A.3}$$

This implies a static revenue function $f(cc, mu, cap, Q, s_0)$ that is known analytically, but can be approximated at a smaller state space in dynamic estimation.

BIBLIOGRAPHY

Abito, J. M. (2017). Agency problems and environmental regulation: Evidence from electric utilities under rate of return regulation. *Working Paper*.

Ackerberg, D., Caves, K., and Frazer, G. (2005). Structural Identification of Production Functions. pages 1–35.

Aden, N., Fridley, D., and Zheng, N. (2009). China's Coal: Demand, Constraints, and Externalities.

Aguirregabiria, V. and Mira, P. (2007). Sequential esitmation of dynamic discrete games. *Econometrica*, 75(1).

Arcidiacono, P. and Miller, R. A. (2013). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 101:1312–44.

Asker, J., Collard-Wexler, A., and De Loecker, J. (2017). Market power, production (mis)allocation and opec. *NBER Working Paper No. 23801*.

Bajari, P., Benkard, C. L., and Levin, J. (2007). Estimating dynamic models of imperfect competition. *Econometrica*, 75(5):1331–1370.

Balat, J. and Sasaki, Y. (2014). Identification and Estimation of Production Functions with Heterogeneous Firms. (1988).

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262.

Borenstein, S., Bushnell, J., and Knittel, C. (1999). Market power in electricity markets: Beyond concentration measures. *The Energy Journal*, 20(4):65–88.

Borenstein, S., Bushnell, J., Knittel, C. R., and Wolfram, C. (2008). Inefficiencies and Market Power in Financial Arbitrage: A Study of California's Electricity Markets. *The Journal of Industrial Economics*, 56(2):347–378.

Borenstein, S., Bushnell, J., and Wolak, F. (2002). Measuring Market Ineffieciencies in California's Restructured Wholesale Electricity Market. *Csem Wp*, 102(June):1–58.

Brandt, L., Biesebroeck, J. V., Wang, L., and Zhang, Y. (2017). WTO Accession and Performance of Chinese Manufacturing Firms. *American Economic Review*, 107(9)).

Brandt, L., Tombe, T., and Zhu, X. (2013). Factor market distortions across time, space and sectors in China. *Review of Economic Dynamics*, 16(1):39–58.

Brandt, L., Van Biesebroeck, J., and Zhang, Y. (2012). Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing. *Journal of Development Economics*, 97(2):339–351.

Brandt, L., Van Biesebroeck, J., and Zhang, Y. (2014). Challenges of working with the Chinese NBS firm-level data. *China Economic Review*, 30:339–352.

Conlon, C. (2012). A dynamic model of prices and margins in the lcd tv industry.

Cooper, R. and Haltiwanger, J. (2006). On the nature of capital adjustment costs. *The Review of Economic Studies*, 73(3):611–633.

Eisenberg, T. (2019). Regulatory distortions and capacity investment: The case of china's coal power industry.

Fabrizio, K. R., Rose, N. L., and Wolfram, C. D. (2007). Do Markets Reduce Costs ? Assessing the Impact of on US Electric Generation Regulatory Restructuring Efficiency. *American Economic Review*, 97(4):1250–1277.

Fackler, P. and Miranda, M. (2002). *Applied Computational Economics and Finance*. MIT Press, 1 edition.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425.

Fowlie, M., Reguant, M., and Ryan, S. (2016). Market-based emissions regulation and industry dynamics. *Journal of Political Economy*, 124(1):249–302.

Gandhi, A., Navarro, S., and Rivers, D. (2013). On the Identification of Production Functions : How Heterogeneous is Productivity ? (May 2006).

Gao, H. and Van Biesebroeck, J. (2014a). Effects of deregulation and vertical unbundling on the performance of China's electricity generation sector. *Journal of Industrial Economics*, 62(1):41–76.

Gao, H. and Van Biesebroeck, J. (2014b). Effects of deregulation and vertical unbundling on the performance of China's electricity generation sector. *Journal of Industrial Economics*, 62(1):41–76.

Hansen, B. (2017). Regression kink with an unknown threshold. *Journal of Business and Economic Statistics*, 35(2):228–240.

Ho, M. S., Wang, Z., and Yu, Z. (2017). China's power generation dispatch. *Resources for the Future*.

Hor, C.-L., Watson, S. J., and Mahjithia, S. (2005). Analyzing the impact of weather variables on monthly electricity demand. *IEEE Transactions on Power Systems*, 20(4):2078–2085.

Hsieh, C.-T. and Klenow, P. (2009). of Economics. *Quarterly Journal of Economics2*, CXII(November):1–55.

Hsieh, C.-T. and Song, M. (2015). [grasp the large, let go of the small: The transformation of the state sector in china].

Ifrach, B. and Weintraub, G. (2017). A framework for dynamic oligopoly in concentrated industries. *Review of Economic Studies*, 84(3):1106–1150.

Jia Barwick, P., Kalouptsidi, M., and Zahur, N. (2019). China's industrial policy: An empirical evaluation. *Cornell Working Paper*.

Kahrl, F., Williams, J., and Hu, J. (2013). The political economy of electricity dispatch reform in china. *Energy Policy*, 53:361–369.

Kalouptsidi, M. (2018). Detection and impact of industrial subsidies: The case of chinese shipbuilding. *Review of Economic Studies*, 85:1111–1158.

Keane, M. and Wolpin, K. (1994). The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence. *Review of Economics and Statistics*, 76(4):1111–1158.

Krasnokutskata, E. and Seim, K. (2011). Bid preference programs and participation in highway procurement auctions. *American Economic Review*, 101:2653–2686.

Lane, N. (2017). Manufacturing revolutions: Industrial policy and networks in south korea. *Working Paper*.

Leslie, P. and Sorensen, A. (2014). Resale and rent-seeking: An application to ticket markets. *Review of Economic Studies*, 81:266–300.

Levinsohn, J. and Petrin, A. (2003). Production Functions Estimating to Control for Using Inputs Unobservables. *Review of Economic Studies*, 70(2):317–341.

Liu, M.-H., Margaritis, D., and Zhang, Y. (2013). Market-driven coal prices and state-administed electricity prices in china. *Energy Economics*, 40:167–175.

Liu, Z. (2013). *Electric Power and Energy in China*. Wiley.

Ma, J. (2011). On-Grid Electricity Tariffs in China: Development, Reform and Prospects. *Energy Policy*, 3:2633–2645.

Marion, J. (2011). Are bid preferences benign? the effect of small business subsidies in highway procurement auctions. *RAND Journal of Economics*, 43(4).

Melitz, M. and Polanec, S. (2013). Dynamic Olley-Pakes Productivity Decomposition with Entry and Exit. *RAND Journal of Economics*, 35(2):362–375.

Newbery, D. M. and Pollitt, M. G. (1997). The Restructuring and Privatisation of Britain's Cegb – Was It Worth It? *Journal of Industrial Economics*, 45(3):269.

Pakes, A. and Olley, S. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6):1263–1297.

Peng, W. (2017). China's silver bullet: Can the transmission grid solve china's problems? *New Security Beat*.

Ren, M., Branstetter, L., Kovak, B., Armanios, D., and Yuan, J. (2019). Why has china overinvested in coal power? *NBER Working Paper 25437*.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11:707–720.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5):999–1033.

Rust, J. and Phelan, C. (1997). How social security and medicare affect retirement in a world of incomplete markets. *Econometrica*, 65(4):781–831.

Ryan, N. (2014). The competitive effect of transmission infrastructure in the indian electricity market. *Working Paper*.

Ryan, S. (2012). The Costs of Environmental Regulation in a Concentrated Industry. *Econometrica*, 80(3):1019–1061.

Saini, V. (2011). Endogenous asymmetry in a dynamic procurement auction. *RAND Journal of Economics*, 43(4).

Scott, P. (2013). Dynamic discrete choice estimation of agricultural land use.

Song, Z., Storesletten, K., and Zilibotti, F. (2011). Growing like China. *American Economic Review*, 101(February):196–233.

Sweeting, A. (2013). Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry. *Econometrica*, 81(5).

Timmins, C. (2002). Measuring the dynamic efficiency costs of regulators' preferences: Municipal water utilities in the arid west. *Econometrica*, 70(2):603–629.

Weintraub, G., Benkard, C. L., and Van Roy, B. (2008). Markov perfect industry dynamics with many firms. *Econometrica*, 76(6):1375–1411.

Weintraub, G., Benkard, L. C., and Van Roy, B. (2017). Computational methods for oblivious equilibrium. *Operations Research*, 58(4.2):1247–1265.

Xu, S. and Chen, W. (2006). The reform of electricity power sector in the pr of china. *Energy Policy*, 34:2455–2465.

Zhang, H. (2015). Non-Neutral Technology, Firm Heterogeneity, and Labor Demand. The University of Hong Kong.

Zhao, X. and Ma, C. (2013). Deregulation, vertical unbundling and the performance of China's large coal-fired power plants. *Energy Economics*, 40:474–483.